Water footprint assessment in space and time to support local and global sustainability

ScuDo

Scuola di Dottorato ~ Doctoral School

WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Environmental Engineering ($XXX$cycle)

# Water footprint assessment in space and time to support local and global sustainability

By


**Marta Tuninetti**

******

**Supervisor(s):**

Prof. Francesco Laio

Prof. Luca Ridolfi

Dr Stefania Tamea


**Doctoral Examination Committee:**

Prof. Matti Kummu, Referee, Aalto University

Dr. Yoshihide Wada, Referee, IIASA

Dr. Carole Dalin, UCL Institute for Sustainable Resources

Prof. Amin A. Elshorbagy, University of Saskatchewan

Dr. Marianela Fader, ICWRGC


Politecnico di Torino

2018

# Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Marta Tuninetti
2018

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

*For Lucia, my beloved grandma*

# Acknowledgements

My doctoral journey has been a *thrilling* experience, surrounded by ideas, people, and motivating challenges.
First of all, I would like to acknowledge my supervisors, Prof. Francesco Laio, Prof. Luca Ridolfi, and Dr. Stefania Tamea, for their precious guidance and support during my PhD.

Thank you, Francesco, for your innovative thoughts, for your trust in myself I could always perceive, for being a continuous source of inspiration. I really admire the way you do Science.

Thank you, Luca, for all your teachings and advices. I sincerely respect and really learned a lot from your depth of knowledge and passion for Science.

Thank you, Stefania, for your constant presence, for your encouragements, for having supported me when I was in London, and for the opportunity to teach in your MSc course.

I am grateful to Prof. Paolo D'Odorico from Berkeley University. As a master student, I had the occasion to visit him at the University of Virginia, where I grew wiser that doing a PhD was the best way to continue *growing*. I would like also to thank Dr Carole Dalin I could join at the University College London for few months to research on water sustainability issues.
Finally, I would like to thank the referees of this thesis, Prof. Matti Kummu and Dr Yoshihide Wada, for their comments and suggestions to improve and strengthen my work.

Special thanks go to all my happy and shiny colleagues I met over the years. I feel very lucky to share this journey (and also some beers) with all of you. Merc*è* *top* Alice, *zia* Irene, and *coinqui* Andrea! Thank you Nicole. Even with an Ocean between us, we ran together in the same direction.

Last, but not least, my gratitude to my *rocky* parents and to my *wolf* brother, Alberto, to my *best* friend, Giulia, and to all my friends *for life*.

*Grazie*, Fabio!

# Abstract

Crop production vastly dominates global freshwater use, accounting for nearly 70% of the total withdrawal and around 90% of the total consumption. Human beings are currently using 30% of precipitation-recharged soil moisture and less than 10% (i.e., 3800 $km^3 \cdot yr^{-1}$) of the maximum available renewable freshwater resources in the word. Notwithstanding, water resource availability is highly variable in space and time, and different studies have shown a significant mismatch between water use and availability. Accordingly, two-third of global population live under conditions of sever water scarcity for at least one month per year. Moreover, as a consequence of larger food demand and changing living standards, toward more caloric and protein intense diets, global water use has increased by 6-8 times during the past century. At the same time, areas equipped for irrigation have doubled with actual irrigation having unavoidable consequences for aquifers and river ecosystems. Future scenarios of climate change are expected to worsen this picture. Indeed, the rising trends of water demand may continue in the future, harshening the conditions in areas reaching critical thresholds of acceptable water balance. In this context, the goals of this thesis are (i) to identify the main determinants of water use efficiency in agriculture; (ii) to introduce a link prediction algorithm applied to the international trade of agricultural goods; (iii) to introduce a novel indicator to monitor the (mis)match between water use and supply. This thesis quantifies the crop water footprint ($CWF$, or amount of water use per unit weight of crop) of nine major crops (i.e., wheat, rice, maize, soybean, barley, potatoes, sugar cane, sugar beet, and cotton) through a daily soil water balance run on a grid with a 5'x5' spatial resolution. The model considers scenarios of rainfed and irrigated crops, also exploring multi-cropping patterns. Quantitative assessments of green and blue (separated into surface and ground) $CWF$ are mapped and analysed in order to identify and monitor the major local

drivers of water use, such as climatic conditions, precipitation rate during the growing season, cropping calendar, soil properties, crop yields and agricultural management practises. Results show that crop yield is the most important determinant of the total $CWF$. Moreover, results of a first-order sensitivity analysis show that, e.g., wheat $CWF$ is mostly sensitive to the length of the growing period, rice $CWF$ to the reference evapotranspiration depth, soybean and maize $CWF$ to the planting date. The $CWF$ model has been adopted also to validate a Fast Track approach, recently developed to study the $CWF$ changes in time, which are generally kept aside in Water Footprint assessments. This approach ascribes the temporal $CWF$ changes only to the yield variations, while it assumes the evapotranspiration depth as time-invariant. This thesis shows the good performance of this approach and also provides an uncertainty analysis. Accordingly, the Fast Track approach shows an error three times smaller than the uncertainty associated with the $CWF$ model. Following the yields patterns, $CWF$ has significantly decreased along the period 1961-2013, but with different rates depending on the crop and the location of the production sites. In the second part of the thesis, the crop water footprint is compared to the local water availability, to assess the sustainability of crop production. In order to understand the size of local (mis)match between crop water use and available water resources, we introduce a water debt repayment time indicator ($WD$). The $WD$ quantifies the time the hydrological cycle takes to replenish the water resources used for annual crop production, distinguishing the different sustainability levels of soil-, surface-, and ground-water. This indicator highlights the locations and typology of threats imposed by agricultural production on water resources. On a global average, we found that wheat and rice production critically overuses ground water resources, especially in China and the US, and cotton production overuses both surface -and ground-water, particularly in the US. Locally, unsustainable annual crop production is found over the Sabarmati basin (due to wheat) in India, and in the Chao Phraya basin (due to rice and sugarcane) in Thailand, where the water debt repayment time exceeds 5 years in many cultivated areas. Including in the same framework analyses on water use efficiencies (through the $CWF$) and measure of water use (un)sustainability (through the $WD$) enables screening analyses at finding specific solutions in cases of low water use efficiencies and/or in critical situation of overuses. While local

drivers monitor the water use for production, global drivers attempt to explore the globalization of water resources that happens through the international trade of agricultural goods. Why do countries become trade patterns, hence establishing a more or less stable relation, which implies externalization of water resources use? The third part of this thesis answers to this question through the elaboration of a threshold-based link prediction algorithm, aiming at finding the drivers behind link activation. Accordingly, a link is expected to exist depending on the predicted virtual water volume traded from the source node to the target node: the link is modelled as active when the volume is higher than 1000 $m^3$/y, non-active otherwise. This algorithm is able to capture 84% of the currently active links and 93% of non-active links. Country population, geographical distance between countries and fertilizers use are the major drivers to explain link existence. The link prediction model may be applied to build future scenarios of virtual water trade, in order to understand how local consumption and production patterns could affect the trade network. In short, the thesis contributes advancing our knowledge in the spatio-temporal explicit water footprint assessments, virtual water trade network, sustainable water use. The models developed in this thesis and the results shown in the following chapters allow (i) to explore pathways toward improved water use efficiencies and more sustainable water withdrawals, (ii) to model backward and forward trade network dynamics, and (iii) to project future water use scenarios.

# Contents

# List of Figures

# List of Tables

# Nomenclature

### Roman Symbols

| | |
|---|---|
| $A$ | adjacency matrix |
| $A$ | harvested area |
| $A^I$ | harvested area in irrigated conditions |
| $A^R$ | harvested area in rainfed conditions |
| $AEI$ | area equipped for irrigation |
| $AP$ | agricultural population |
| $AV$ | agricultural value |
| $AWC$ | available water content |
| $b$ | binary matrix |
| $CWF$ | crop water footprint |
| $CWF_b$ | blue crop water footprint |
| $CWF_{gw}$ | ground crop water footprint |
| $CWF_g$ | green crop water footprint |
| $CWF_{sw}$ | surface crop water footprint |
| $D$ | geographic distance between countries |
| $D_{ev}$ | root zone depletion in the evening |

| | |
|---|---|
| $D_{mo}$ | root zone depletion in the morning |
| $ET_0$ | reference evapotranspiration |
| $ET_b$ | blue evapotranspiration |
| $ET_g$ | green evapotranspiration |
| $ET_a$ | crop actual evapotranspiration |
| $F$ | bilateral trade matrix of virtual water flux |
| $f$ | time-averaged country yield |
| $FPR$ | false positive rate |
| $g$ | ratio of cell yield to country yield |
| $GDP$ | gross domestic product |
| $I$ | irrigation volume |
| $K$ | potash fertilizer in tonne of nutrients |
| $k_s$ | water stress coefficient |
| $K_y$ | yield response factor |
| $k_c$ | crop coefficient |
| $K_{imp}$ | in-node degree |
| $K_{out}$ | out-node degree |
| $LGP$ | length of the growing period |
| $N$ | nitrogen fertilizer in tonne of nutrients |
| $PD$ | planting date |
| $Ph$ | phosphate fertilizer in tonne of nutrients |
| $PR$ | crop production in tonnes |
| $Pr$ | precipitation |

$R_{s,l}$ annual renewability rate of source $s$ at location $l$

$RAW$ readily available water

$rta$ regional trade agreement

$SI$ sensitivity index

$TAW$ total available water

$TPR$ true positive rate

$VWT$ virtual water trade

$VWTN$ virtual water trade network

$vwv$ virtual water value

$W_{cr}$ global trade of single crop expressed in tonne

$WD$ water debt indicator

$wd$ per-capita demand of water embedded in agricultural goods

$WF$ water footprint of total production

$wp$ virtual water of per-capita agricultural production

$x_{ref}$ reference parameter for the sensitivity analysis

$Y_a$ crop actual yield

$Zr$ crop rooting depth

P population

**Greek Symbols**

$\alpha^{cl}$ yield climatic factor

$\alpha^{man}$ yield management factor

$\epsilon$ $CWF$ model error

$\epsilon'$ FT approach error

$\rho$ soil water depletion factor

$\theta_{FC}$ soil water content at field capacity

$\theta_{WP}$ soil water content at wilting point

# Chapter 1

# Introduction

Food production is inextricably linked and reliant upon freshwater resources: in fact, the vast majority of global freshwater use (nearly 70% of the total withdrawal [49] and around 90% of the total consumption [73]) is devoted to the production of agricultural commodities [47, 100], largely for human consumption, and water is a major factor controlling food availability [61, 133]. Rain-fed agriculture, sustained by precipitation-recharged soil moisture (the so-called green water [44, 48]), covers 80% of the cultivated land worldwide. On the remaining 20% of cultivated land, irrigated agriculture provides 42% of global food production [49]. At present, nearly 30% of green water resources and only 10% of maximum available blue water (i.e., water withdrawal from surface- and ground-water bodies) are used [101]. Figure 1.1 shows the global map of blue water consumption from agricultural, industrial, and domestic sectors; the total volume is 2000 $km^3 \cdot yr^{-1}$ according to [144]. It is evident that water consumption is unevenly distributed worldwide and the largest consumption occurs in India, Pakistan, China, USA, and Mexico.

Global food demand and rising living standard increased the global water use by 6-8 times from 1900 to 2010 [43, 142], highlighting the growing importance of each drop of water, as water consumption gets closer to water availability [146]. Areas equipped for irrigation have doubled in the past 50 years and are projected to increase by 6% by 2050, with an increase of 38% of irrigated food production [49]. Withdrawals from surface- and ground- water bodies is intensifying, especially during precipitation shortages. Intensification of irrigation demand has already led to alarming levels of water depletion in

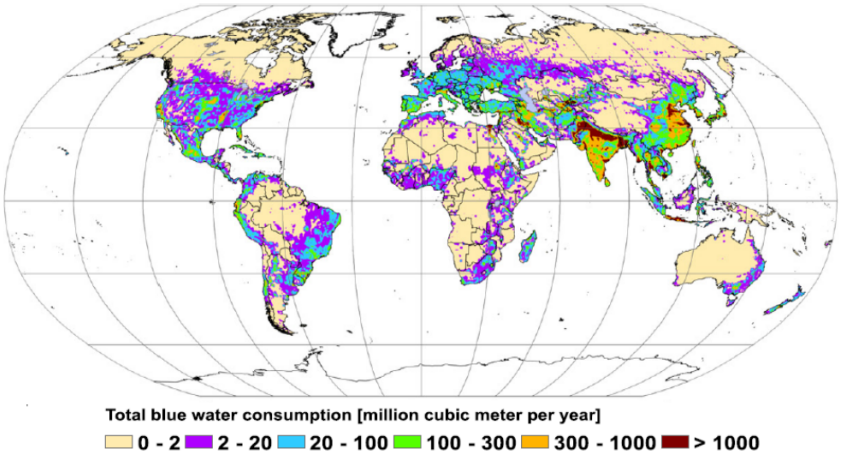


Fig. 1.1 Total blue water consumption for year 2010, expressed in million cubic meter per year. Source: *Wada et al.* [144].

important aquifers [59, 33] and river systems [112, 143] worldwide. Moreover, the world population of 7.2 billion in early 2015 is projected to increase to 9.6 billion in 2050 [139], resulting in a 70% increase in the global demand for agricultural products [49]. Climate change is expected to worsen this picture by increasing the spatial heterogeneity of water resource availability [49]. Similarly to water consumption, also water availability is highly heterogeneous in space (with two billion people currently living in highly water stressed areas [4, 46, 83]), and time, due to evolving climate patterns [101, 3]. In a recent study, *Mekonnen et al.* [93] shown global water scarcity- calculated as the ratio between net water withdrawal and blue water availability- on a monthly basis. The map in Figure 1.2 points out how water scarcity is often a major issue in area with high population density or intense irrigated agriculture. Scientists, policy makers, and the general public are realizing that meeting the competing water needs of ecosystems and societies is a major challenge for this century [40, 64]. It is urgent to reduce and reverse depletion trends to ensure a sustainable agricultural water use, i.e. a use that does not compromise the Earth's water resources for future generations [21].

In order to explore the nexus between food production and water consumption, *Hoekstra et al.* [67, 75, 69] introduced the Water Footprint ($WF$) as an indicator of water use related to goods and services produced or consumed by an individual (or a country), separating green water from blue water. The notion of water footprint is tightly connected to that of virtual water content,

which represents the amount of water that is conceptually embedded (though not physically present) in a good [6]. The concept of virtual water content has been used for the first time by Allan [7], who suggested that virtual water import, i.e., the water embedded in imported goods, was a mechanism that contributed to compensate for water shortage in Middle East countries.
In light of the fact that agriculture is the major water-consuming sector, many studies have focused on water footprint of crops. In particular, they focused on the efficiency of water use, expressed by the crop water footprint ($CWF$) quantified as the volume of water evapotranspired during the growing season divided by the crop yield [89, 5]. In recent years, there have been various attempts to assess global water use for crop production at different geographic scales, generally averaged over a specific time period. Table 1.1 summarises the main studies on the global $VWC$ assessment.

Table 1.1 Studies about the global $VWC$ of crop production at different geographic scales.

| Study | Scale | Resolution | Period | Crop yield | Rainfall | $ET_0$ | Soil | Sensitivity |
|---|---|---|---|---|---|---|---|---|
| *Rost et al.* [114] | global | 0.5 deg | 1971-2000 | - | 0.5 deg | 0.5 deg | 0.5 deg | NO |
| *Hanasaki et al.* [62] | global | 0.5 deg | 1985-1999 | country | 1 deg | 1 deg | uniform | NO |
| *Liu et al.* [85] | global | 0.5 deg | 2000 | - | 0.5 deg | 0.5 deg | 5 arcmin | NO |
| *Siebert et al.* [126] | global | 5 arcmin | 1998-2002 | 5 arcmin | 10 arcmin | 10 arcmin | 5 arcmin | NO |
| *Mekonnen et al.* [89] | global | 5 arcmin | 1996-2005 | 5 arcmin | 0.5 deg | 10 arcmin | 5 arcmin | NO |
| *Zhuo et al.* [153] | local | 5 arcmin | 1996-2005 | 5 arcmin | 0.5 deg | 10 arcmin | 5 arcmin | YES |
| *Bocchiola et al.* [16] | local | 5 arcmin | 2001-2010 | 5 arcmin | 0.5 deg | 10 arcmin | 5 arcmin | YES |

*Rost et al.* [114] mapped (with a spatial resolution of 0.5 by 0.5 arc degree resolution) the total consumption of green and blue water by rainfed and irrigated agriculture for 12 crops and 9 plant functional types, for a reference period 1971-2000. *Hanasaki et al.* [62] reported the global green and blue water consumption for five crops and three livestock products at 0.5 by 0.5 arc degree resolution, along the period 1985-1999. *Liu et al.* [85] made a global estimate of green and blue water consumption for crop production at 0.5 arc degree resolution. *Siebert et al.* [126] quantified the green and blue virtual water content of 26 crops with a spatial resolution of 5 by 5 arc minute in the period 1998-2002. In a recent study, *Mekonnen et al.* [89] estimated the green, blue, and grey water footprints of 126 crops with a spatial-resolution of 5 by 5 arc minute, relative to the period 1996-2005. Overall, it has been shown that crop water footprint is highly heterogeneous in space due, e.g., to different climate and soil conditions, fertilizer application rates, and agricultural mechanization

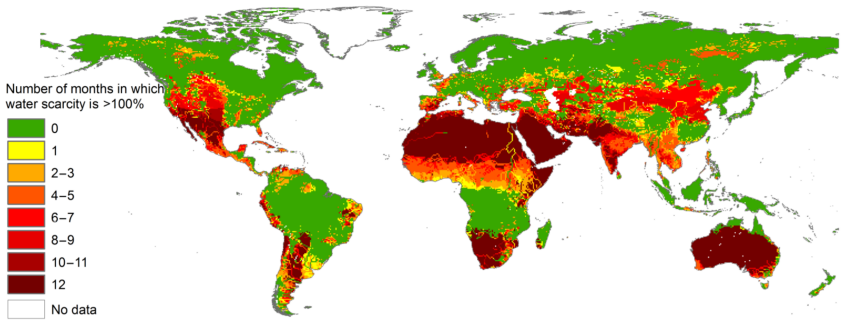


Fig. 1.2 Number of months per year in which water scarcity exceeds 1 Source: [93].

level, even at the sub-national level. All these studies depend on a large set of assumptions about the modelling structure, input parameters, and datasets used. Only very few studies, however, have developed a sensitivity analysis of the water footprint estimates to define the accuracy of the final outcomes. Their focus has been on specific regions, such as the Yellow River basin in China [153] and the Po valley in Northern Italy [16]. To date, a global-scale sensitivity analysis with high spatial-resolution is still missing. Filling this gap is one of the goals of the present thesis, through (i) the development of a high-resolution model to quantify crop water footprint worldwide and (ii) the performance of a first-order sensitivity analysis, to understand how data uncertainty propagates through the crop water footprint estimates and to identify the model inputs that significantly affect the model outputs.

While a great deal of attention has been devoted to the $CWF$ variability in space, less attention has been paid to its variability in time even though climatic fluctuations and yield variations have been remarkable in the past decades [98, 134, 111]. To date, only local studies have evaluated a time-varying crop water footprint [128, 149, 109], with particular regard to the Chinese case [129, 155, 154]. Some studies investigating the time variability of the virtual water embedded in trade used annual trade data of agricultural goods, i.e., available on the FAOSTAT database, and time-averaged crop water footprint as provided by *Mekonnen et al.* [89]. At the same time, also some local studies dealt with the time variability of the VWT with constant virtual water content [78, 121].
However, considering a constant water footprint precludes analyses on the implications of climate patterns and yield trends on the virtual water content and, thus, on the virtual water trade. In order to keep pace with this issue, a number

of studies have adopted a simple approach that ascribes the time variability of virtual water content only to yield trends, leaving out the effects of evapotranspiration variations [81, 80, 31, 32]. Studies have adopted this approach both for global [118, 24, 39] and local [30, 79] water footprint assessments, but the feasibility of this approach has yet to be proved. Can this approach capture the main $CWF$ temporal variability? How big is the error arising from the assumption of constant evapotranspiration? What is the effect of $CWF$ variability on the virtual water trade? The present thesis addresses these questions by (i) providing a systematic validation of the method (here referred as the Fast Track method), (ii) providing a comprehensive assessment of the associated uncertainty, and (iii) giving an example of application to highlight its suitability.

As mentioned above, in a framework where water resources suffer from increasing pressures from population growth, economic development and climate change [142, 60], the international food trade is vital for food security [40, 64]. Through the international trade of agricultural goods, water resources that are physically used in the country of production are 'virtually' transferred to the country of consumption. Virtual water trade has often been recognized for its ability to improve physical and economic access to food commodities in water scarce regions, allowing nations to save domestic water resources through the import of water-intensive products [26, 150]. Thus, food trade leads to a global redistribution of freshwater resources [67], although it is recognized that commodities are being traded, and not water.
The virtual water trade between regions has been investigated by a remarkable variety of scientific contributions. A larger number of studies investigated the temporal evolution and dynamics of the virtual water trade (VWT) associated to the international trade of agricultural goods reference therein [10]. The VWT has been recognized for its ability to improve access to water for food production in those countries where water scarcity is a major concerning issue, i.e. through the import of water-intensive products [27, 64]. It has been shown how both the virtual water volume embedded in internationally-traded goods and the number of trade relations grew significantly between 1986 and 2010 [22, 23], mainly driven by population, GDP, and geographical distances between countries [132, 137]. Some authors (e.g., [71, 100, 72, 27, 62]) evaluated the country virtual water balance in relation to national water needs and water

availability. Their studies revealed that virtual water is exchanged primarily from countries of high crop water productivity to countries of low crop water productivity, generating a global saving in water use [150].
Recently, the analyses of virtual water trade by tools of complex network theory [99, 15] have also attracted much attention. The idea has been to depict the virtual water trade as a network where countries play the role of nodes and links describe the import/export relations between any two countries (i.e., virtual water trade network, $VWTN$). Important insights about the global architecture of the $VWTN$ have been provided by *Konar et al.* [81]: their (data-based) analyses quantified the topology of international trade, providing evidence for the existence of the weighted rich club phenomenon and uncovering a global trading hierarchy in which dominant nations connect most peripheral portions of the network. *Suweis et al.* [130] developed a simple model that captures the key features of the network by assuming as sole controls each country's gross domestic product and yearly rainfall on agricultural areas, *Tamea et al.* [132] identified population, gross domestic product, and geographical distance of countries as major drivers of virtual water fluxes, whereas *Fracasso* [55] found that also water endowments and water pressure in terms of water withdrawal impact the bilateral import and export of virtual water.
Other contributions have investigated the time evolution of the virtual water trade network. Such studies highlighted the dynamical and intermittent behaviour of the network where a number of links are created and dismissed every year [22, 23]. *Dalin et al.* [31] showed that the $VWTN$ has grown significantly between 1986-2007, with nodes becoming more interconnected and exchanging larger virtual water volumes, although the distributions of the main network statistics have remained stable, allowing predictive models of the network statistical characteristics to be developed. Also the fitness model developed by *Dalin et al.* [32] showed good performances in reproducing the structural properties of the $VWTN$, enabling the forecasting of the structure of the network under future political, economic and climatic scenarios.
In spite of the growing efforts devoted to unfold the virtual water network structure and dynamics through its general features, a poorly explored problem is to understand the single association between any pair of countries involved in the international trade: this can be formalized as a *link prediction task*. Link prediction is the problem of predicting the existence of a link between two

nodes, based on the attributes of nodes and/or on the network topology [58]. In the context of network theory, link prediction has attracted much attention from different research communities and several algorithms are available. The main-streaming class of algorithms are similarity-based algorithms [87], where similarity can be defined by using the attributes of nodes; however, the attributes of nodes are generally hidden, and thus, these algorithms are usually based solely on the network structure or topology (i.e., structural-similarity indices). In the framework of complex network, link prediction is generally used to extract missing information, identify spurious interactions, and evaluate evolving network mechanisms starting from a given network which is usually undirected and unweighted [87]. Conversely, the specific problem addressed in this thesis is the link prediction between any two nodes over a directed and weighted network, without any knowledge of the network topology, but only knowing the country attributes (e.g., population, gross domestic product, water demand) and the link characteristics (e.g., geographical distance). Specifically, the network topology is assessed by a threshold algorithm that establishes the link existence on the basis of the possible weight of the link itself estimated by means of the gravity law model: i.e., the link exists if the expected traded flow is greater than 1000 $m^3$ of virtual water.

A critique that has been made to virtual water studies is that they do not account for water endowments of countries, thus ignoring the impact of food trade (and production) on local resources. An important development of the water footprint indicator, necessary to assess the impact of food production on water resources across products and regions, is to relate the crop water use to locally available renewable water resources. In the past decades, many water indices have been developed [4, 2, 147, 144], generally comparing water use with local availability and denoting as "highly water stressed" those areas where the ratio is above 0.4. Recently, also the water footprint has been analysed in the sustainability framework, showing e.g., where blue $WF$ exceeds maximum sustainable levels at the monthly scale [74]. *Yano et al.* [151] proposed a water scarcity footprint, scaling the $WF$ by a factor expressing local water unavailabilities, as an indicator of water use impacts [152]. *Brauman et al.* [19] recently provided a comprehensive picture of the areas vulnerable to water shortage by the introduction of a water depletion metric. However, although the

need to compare water use to water availability has been acknowledged and these indicators are useful to identify areas of stress and water resource depletion, they only provide broad pictures of agriculture sustainability [70]. In fact, indicators generally consider all crops together, not allowing the identification of critical crops, which may help to define alternative strategies or to track unsustainable trade flows. In addition, most indicators do not separate the use of surface water from groundwater use although this distinction is fundamental because of the different renewability of such resources. and do not consider the role of green water scarcity in sustainability assessments [45]. Moreover, some weighted water footprint metrics often lack a physical interpretation, although proving to be useful for country comparisons, and are less effective for critical analyses [68]. An aim of this thesis is to fill these gaps in the evaluation of agricultural water use sustainability, separating irrigation from surface- and ground-water, accounting for green water use and considering crops separately. In order to measure the sustainability of agricultural production with respect to (local) water resources availability, the concept of "Water Debt" ($WD$) is introduced. The term, borrowed from the carbon footprint literature [54] and inspired by early considerations about sustainability [66], is used to indicate the payback time required by the hydrological cycle to replenish or refresh the water resources (i.e., soil moisture, surface water bodies, and aquifers) used for annual crop production. The $WD$ enables to compare the sustainability of crop water use across crops and regions by including both the local water use and the renewable availability. It allows one to identify most threatened water sources and helps decision- and policy-makers to prioritize water saving strategies (e.g., increasing water use efficiency) and plan sustainable withdrawals and supply of freshwater (both green and blue).

The overall thesis has several, but tightly connected, objectives as shown in Figure 1.3. Motivated by the uneven spatio-temporal distribution of water resources and considering the issue of balancing water demand with water availability, the present study asses (both locally and globally) the main drivers of water agricultural water use and the water resources globalization connected to the international trade of agricultural goods. Moreover, the thesis tries to assess the impact of water use on local resources by means of a newly developed indicator.
The thesis has been designed into four sub-research projects, which will be

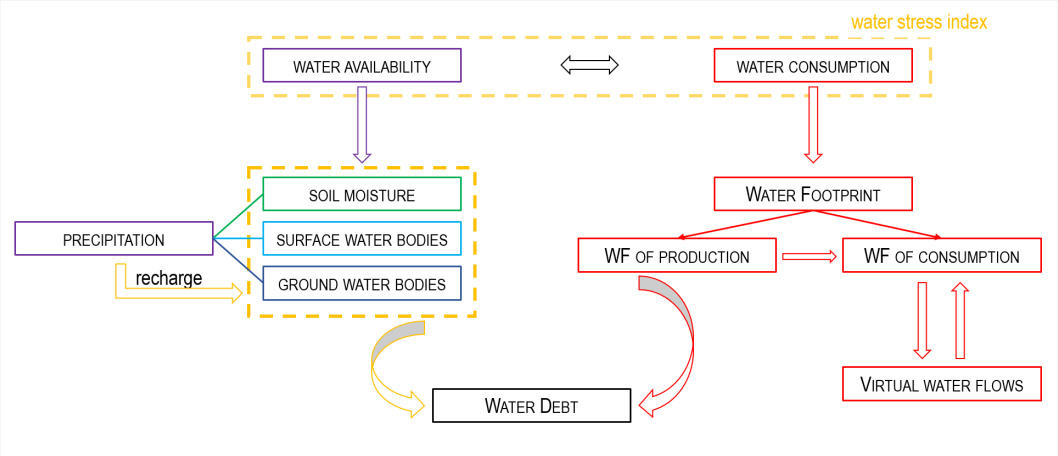


Fig. 1.3 Map of the main key concepts of this thesis and their relationship.

described in Chapters 2-5.

Chapter 2 assesses the water use geography of four crops- wheat, rice, maize, and soybean- over 1996-2005. The crop water use is quantified through a daily soil water balance, under rainfed and irrigated conditions. High-resolution data, such as precipitation, evapotranspiration, soil properties, are used as input parameters to the model and their associated uncertainties is explored through a first order sensitivity analysis. Results are shown both as maps and as regional/global averages to convey at the same time local and global information.

Chapter 3 deals with the temporal dimension of crop water footprint over the years 1961-2014. A Fast-Track method, which ascribes water footprint variability only to yield changes, is described and validated, and an uncertainty analysis is performed. The Fast-Track method is then applied to the case of virtual water trade to show its significance and to foster its adoption in analyses where the temporal dimension of water footprint is relevant.

Chapter 4 introduces a novel link-prediction algorithm to investigate the dynamical behaviour of the virtual water network associated with the international trade of over 300 products. The algorithm is based on the country characteristics and the link properties, while it does not require the network topology as mainstream link prediction algorithms do.

Chapter 5 focuses on the sustainability of water use in agriculture. To deal with water use sustainability, the water uses computed in Chapter 2 are extended to other five crops, namely barley, potatoes, sugar beet, sugar cane, and cotton,

and the blue water use is split into surface and ground water use. The concept of Water Debt is introduced as a measure of the impact of crop water use on renewable resources.
Finally, Chapter 6 concludes the thesis with a synthesis of the achievements and puts the results into future research perspectives, suggesting possible directions and providing solutions to overcome the intrinsic limitations of the current studies.

# Chapter 2

# The water footprint of crop production worldwide

*The work described in this chapter has been partially derived from paper [136].*
Most of the human appropriation of freshwater resources is for agriculture. Water availability is a major constraint to mankind's ability to produce food. The notion of water footprint ($WF$) provides an effective tool to investigate the linkage between food and water resources as a function of climate, soil and agricultural practices.
Generally, the water footprint ($WF$) of a good is the amount of water needed to produce the good along the whole production chain [6], in primary crops most water use is related to the evapotranspiration process. When applied to agricultural goods, water footprint is often evaluated per unit of product/crop in order to provide a measure of the efficiency of water use. The unit water footprint (or crop water footprint, $CWF$) allows for a comparison of different commodities on the basis of their water use; hence, the concept kindles analyses about the water use efficiency and how to improve it. Moreover, the crop water footprint can be quantified in terms of green and blue water components, depending on whether the water is contributed by precipitation water stored in the (top of) soil and vegetation, or by surface and groundwater used for irrigation and food processing. A number of studies has explored the spatial variability of the crop water footprint, at different spatial scales, for a wide basket of products; however, shortcomings can be found. In fact, only few studies assessed the $CWF$ at high spatial resolution worldwide [89, 126, 41] and

only the study by Siebert et al. [126] considered multi-cropping periods along the year, which are often used in crop production. Moreover, only local studies [153, 16] performed a sensitivity analysis of the crop water footprint to major input parameters. These main shortcomings in the literature are addressed in this Chapter. The spatial variability of crop water footprint is explored in the first part of the Chapter, disentangling its dependency on climate and crop yields. To this end, we calculate the crop water footprint of four crops -wheat, rice, maize, soybean-for the entire world developing a high-resolution (5 by 5 arc minute) soil water balance model. In the second part of the Chapter, a sensitivity analysis is performed to study how the unavoidable uncertainties existing in the (spatial) climate and soil data affect the calculation of the crop water footprint.

Results suggest that food production almost entirely depends on green water (>80%), but, when applied, irrigation makes crop production more water efficient, thus requiring less water. The spatial variability of the $CWF$ is mostly controlled by the spatial patterns of crop yields with an average correlation coefficient of 0.83. The results of the sensitivity analysis show that wheat is most sensitive to the length of the growing period, rice to reference evapotranspiration, maize and soybean to the crop planting date. The $CWF$ sensitivity varies not only among crops, but also across the harvested areas of the world, even at the sub-national scale.

## 2.1 Estimating the green and blue water footprints in crop production

The crop water footprint, $CWF$, is evaluated globally at the spatial resolution of 5 by 5 arc minutes, corresponding to pixels of about 9 km by 9 km at the equator. We consider both rainfed and irrigated production conditions, as well as multi-cropping practices (i.e., a crop can be grown on the same land more than once a year). $CWF$ estimates are referred to the time interval from 1996 to 2005. We chose a data range of 10 years in order to have input data independent of inter-annual fluctuations and typical of each grid cell; more specifically, we build our reference period centred on year 2000 because this is the most frequent reference year in the global agricultural datasets used in this

study (e.g., crop calendar, crop yields, harvested areas).
$CWF$ is defined in each pixel as the ratio between the water evapotranspired by the crop during the growing seasons of a year $y$, $ET_{a,y}$ (mm), and the crop actual yield, $Y_a$ (ton ha$^{-1}$), as

$$CWF = \frac{10 \cdot ET_{a,y}}{Y_a} \qquad \left[\frac{\mathrm{m}^3}{\mathrm{ton}}\right], \tag{2.1}$$

where the factor 10 converts the evapotranspired water height expressed in mm into a water volume per land surface expressed in m$^3$ ha$^{-1}$.
In regions where more than one crop per year is planted and harvested (i.e., there are multiple growing seasons), the actual evapotranspiration of a year, $ET_{a,y}$, is calculated as the weighted average (with respect to the area $A_n$ cultivated during the growing period $n$ with $n = 1, 2, ...$) of the total actual evapotranspiration $ET_{a,LPG,n}$ (mm) of each growing season, as

$$ET_{a,y} = \frac{\sum_n (ET_{a,LGP,n} \cdot A_n)}{\sum_n A_n} \qquad [\mathrm{mm}], \tag{2.2}$$

where $LGP$ is the length of each growing period. Depending on agricultural practices, climate and soil properties, the crop evapotranspires green ($ET_{g,y}$) and/or blue water ($ET_{b,y}$). Thus, the total water evapotranspired by the crop during the growing seasons of a year can be written as the sum of a green and a blue component,

$$ET_{a,y} = ET_{g,y} + ET_{b,y}. \tag{2.3}$$

### 2.1.1 Crop evapotranspiration over a single growing season

The total water evapotranspired by the crop in a single growing season, $ET_{a,LGP}$ (mm), is obtained by summing up over the length of the growing period ($LGP$) the daily actual evapotranspiration, $ET_{a,j}$ (mm day$^{-1}$), i.e.,

$$ET_{a,LGP} = \sum_{j=1}^{LGP} ET_{a,j} \qquad [\mathrm{mm}], \tag{2.4}$$

with $j$ indicating the day of the growing period. $LGP$ is delimited by the planting ($PD$) and harvesting dates taken from [108]. This data set distinguishes between rainfed and irrigated production and provides the month in which the growing period starts and ends at 5 by 5 arc minute resolution, considering multi-cropping practices, for year 2000. We initially assume that the cropping period starts and ends in the middle of the month.

Daily crop evapotranspiration, $ET_{a,j}$, is calculated following [8], a well-established approach for the virtual water content assessment [89, 126, 153]. $ET_{a,j}$ is defined as

$$ET_{a,j} = k_{c,j} \cdot ET_{0,j} \cdot k_{s,j} \qquad \left[\frac{\text{mm}}{\text{day}}\right], \tag{2.5}$$

where $k_{c,j}$ is the daily crop coefficient, $ET_{0,j}$ is the daily reference evapotranspiration (mm day$^{-1}$) from a hypothetical well-watered grass surface with fixed crop height, albedo and canopy resistance, and $k_{s,j}$ is the daily water stress coefficient depending on the available soil water content, with a value between 0 (maximum water stress) and 1 (no water stress).

The crop coefficient, $k_{c,j}$, depends on crop characteristics and, to a limited extent, on climate. It is influenced by crop height, albedo, canopy resistance and evaporation from bare soil. During the growing period, $k_{c,j}$ varies with a characteristic shape divided into 4 growing stages (I: initial phase, II: development stage, III: mid-season, IV: late season) of $l_I$, $l_{II}$, $l_{III}$, and $l_{IV}$ days length, respectively, that reads

$$k_{c,j} = \begin{cases} k_{c,in} & \text{j}\in \text{ I stage} \\ j \cdot \frac{k_{c,mid}-k_{c,in}}{j-l_I} & \text{j}\in \text{ II stage} \\ k_{c,mid} & \text{j}\in \text{ III stage} \\ j \cdot \frac{k_{c,f}-k_{c,mid}}{j-l_I-l_{II}-l_{III}} & \text{j}\in \text{ IV stage.} \end{cases} \tag{2.6}$$

We use values from [8] for the constants $k_{c,in}$, $k_{c,mid}$, $k_{c,f}$, while the length of each stage, $l_{st}$, is calculated as a fraction, $p_{st}$, of the length of the growing period ($l_{st} = p_{st} \cdot LGP$); $p_{st}$ is defined for each stage (with $st = I - IV$) according to [89], whose study provides specific values of $p_{st}$ for different climatic regions.

Lengths are rounded to the nearest integer and the stage I is adjusted to guarantee the exact length of the growing period.

Monthly long-term average reference evapotranspiration data, $ET_{0,m}$, at 10 by 10 arc minute resolution are given by [52]. These data are converted to 5 by 5 arc minute data by subdividing each grid cell into 4 square elements and assigning them the correspondent 10 by 10 values. Daily $ET_{0,j}$ values are determined through a linear interpolation of monthly climatic data and attributing the monthly $ET_{0,m}$ value to the middle of the month. For sake of simplicity, we consider months 30 days long. These conversions introduce uncertainties in the calculation of the $CWF$, but they are necessary because of the lack of daily evapotranspiration data at 5 by 5 arc minute resolution.

The water stress coefficient typical of the cell, $k_{s,j}$, varies during the growing period depending on the total available water content ($TAW$) and the readily available water content ($RAW$) in the root zone [8]. The water stress coefficient is evaluated considering two different types of production: rainfed production ($R$), in which crops are fed only by, and irrigated production ($I$), in which crops are irrigated when necessary in order to prevent the emergence of water stress. In the irrigated production the water stress coefficient, $k_{s,j}$, is equal to 1 throughout the growth period. In the rainfed production, the computation of the $k_{s,j}$ daily value is detailed in Section 2.1.2. In its evaluation, we use for the first time in a $CWF$ assessment (to the best of our knowledge) the 30 arc second maps of the available water content ($AWC$) given by [53] and the 10 arc minute maps of monthly precipitation given by [98]. Since the daily $k_{s,j}$ is different in the two production types (rainfed vs irrigated), as well as the $ET_{0,j}$ (i.e., the growing period can have different planting dates in rainfed and irrigated conditions), the daily actual evapotranspiration (green+blue), $ET_{a,j}$, calculated with equation (2.5), is different in the two production types. The green component for rainfed crops, $ET^R_{g,j}$, is equal to the total evapotranspiration, $ET^R_{a,j}$. The blue component for irrigated crops, $ET^I_{b,j}$ (notice that by definition, $ET^R_{b,j} = 0$), is obtained as the amount of irrigation water provided to the crop; the green component is the difference between the $ET^I_{a,j}$ and $ET^I_{b,j}$ values. The total, green and blue evapotranspiration over the growing period are given by equation (2.4), both for rainfed ($ET_{g,LGP}$) and irrigated conditions ($ET^I_{g,LGP}$ and $ET^I_{b,LGP}$).
The overall evapotranspiration of green and blue water from the cell, $ET_{g,LGP}$

and $ET_{b,LGP}$, is the weighted mean of the rainfed and irrigated evapotranspiration,

$$ET_{g,LGP} = \frac{ET^R_{g,LGP} \cdot A^R + ET^I_{g,LGP} \cdot A^I}{A^R + A^I}, \tag{2.7}$$

$$ET_{b,LGP} = \frac{ET^I_{b,LGP} \cdot A^I}{A^R + A^I}, \tag{2.8}$$

where weights, $A^R$ and $A^I$, are the harvested areas given by [108]. This data set distinguishes rainfed and irrigated production, providing the harvested areas of 26 main crops, for each growing season. The procedure to evaluate the values of $ET_{g,LGP}$ and $ET_{b,LGP}$ is repeated for each growing season of a year; equation (2.2) is then applied to determine the green and blue evapotranspiration of a year.

Due to peculiarities of the rice cultivation, the $CWF$ estimates need further details. Rice is typically cultivated in wetland or upland systems. About 85% of the rice in the world is grown in wetland systems and about 75% of rice production is obtained from irrigated sites [18]. In wetland rice cultivations, paddy fields are prepared and the soil is kept saturated. Basically, in the month before sowing or transplanting, water is used to saturate the root zone and the amount of water needed depends on the soil type and rooting depth, we considered a volume per unit area of 200 mm, as suggested by [18]. Moreover, during the growing season a constant percolation of water occurs below the root zone, whose rate is affected by a number of soil factors [148]. In this study, we assume a 2.5 mm day$^{-1}$ flux, corresponding to rather impermeable soils with a clayey texture [25].

### 2.1.2 Computation of the water stress coefficient

The water stress coefficient describes the effect of soil water shortage on crop evapotranspiration. *Allet et al.* (1998) [8] defined the daily water stress coefficient as

$$k_{s,j} = \frac{TAW_j - D_{mo,j}}{TAW_j - RAW_j} \qquad [-], \tag{2.9}$$

where $TAW_j$ (mm) is the total available water content in the root zone, $RAW_j$ (mm) is the readily available water content, and $D_{mo,j}$ (mm) is the root zone depletion in the morning (i.e., the water shortage relative to field capacity).

$TAW_j$ depends on the available soil water content per meter depth, $AWC$ (mm $m^{-1}$), and on the daily rooting depth, $Zr_j$ (m) , according to

$$TAW_j = AWC \cdot Zr_j \qquad [\text{mm}]. \tag{2.10}$$

Grid based data on $AWC$, at 30 by 30 arc second resolution, were taken from [53] and converted to 5 by 5 arc minutes through an average. The available water content refers to the capacity of the soil to retain water available to plants. It is equal to the difference between the soil water content at field capacity ($\theta_{FC}$) and the water content at wilting point ($\theta_{WP}$,i.e., the point at which plant will permanently wilt) as shown in Figure 2.1. The rooting depth, $Zr_j$, is given by [8]; this value generally increases during the first two growing stages up to a maximum value (dependent on crop type and irrigation conditions) and then it remains constant until the harvest day.

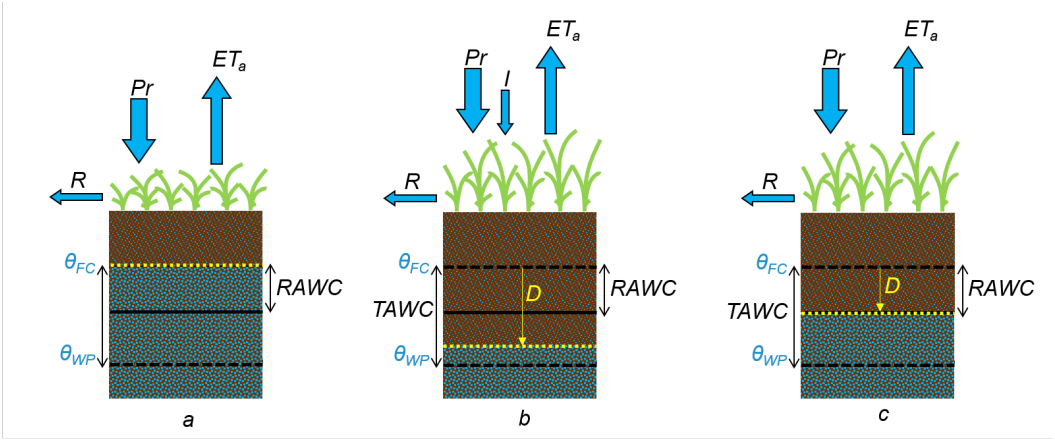


Fig. 2.1 Water balance of the root zone along three significant days of the growing period: the planting date (a), a typical day under water stress condition (b), and a typical day after applying irrigation.

$RAW_j$ is the water that crops can use for evapotranspiration before water stress and stomata closure begin. It is given by

$$RAW_j = \rho \cdot TAW_j \qquad [\text{mm}], \tag{2.11}$$

where, $\rho$ is the soil water depletion factor that can be depleted from the root zone before moisture stress occurs, and it is different among species. We assumed $\rho$ to be constant during the growing season ($\rho$ values are given by [8]). The different values of $TAW$ and $RAW$ in rainfed and irrigated production are due to the different rooting depth (which is deeper in rainfed production). Root zone depletion is recorded in the morning, $D_{mo,j}$, depending on daily precipitation, irrigation, and crop evapotranspiration.
In rainfed production ($R$), the root zone depletion in the morning, $D_{mo,j}$, is equal to the one recorded at the end of the previous day ($D_{ev,(j-1)}$), minus the daily precipitation value, $Pr_j$,

$$D_{mo,j} = D_{ev,(j-1)} - Pr_j \qquad \left[\frac{\text{mm}}{\text{day}}\right]. \tag{2.12}$$

Daily precipitation is obtained equally distributing the monthly climatic precipitation along the growing period with daily frequency. For sake of simplicity, all months are assumed 30 days long. Monthly climatic precipitation are available in the literature at 10 by 10 arc minute resolution [98]; we convert this data to 5 by 5 arc minute grid cells as done in Section 2.1.1 for reference evapotranspiration. $D_{mo,j}$ is equal to 0 on the planting day (see Figure 2.1a). In the evening $D_{mo,j}$ increases because of crop evapotranspiration, as

$$ET_{a,j} = k_{c,j} \cdot ET_{0,j} \cdot \frac{TAW_j - D_{mo,j}}{TAW_j - RAW_j}, \tag{2.13}$$

$$D_{ev,j} = D_{mo,j} + ET_{a,j} \qquad \left[\frac{\text{mm}}{\text{day}}\right]. \tag{2.14}$$

We did not consider water lost by deep percolation, and the capillary rise was assumed equal to zero, whereas water excess (leading to negative values of $D_{mo,j}$) were cut off at zero and the exceeding precipitation was assumed to be lost as surface runoff. In rainfed conditions, the water volume evapotranspired by the crop during the growth period is totally green, $ET_{g,LGP} = ET_{a,LGP}$. Generally, during the first two stages of the growing season the crop has sufficient water to evapotranspire, i.e., $D_{mo,j} < RAW_j$. Conversely, at the

beginning of the third stage (when the crop water requirement is higher) water stress conditions are more frequent ($D_{mo,j} \geq RAW_j$, Figure 2.1b) and the soil water content gets closer to the wilting point. In the rainfed scenario only precipitation can increase the water content up to field capacity.

In irrigated production ($I$), irrigation is required when rainfall is insufficient to compensate for the water loss by evapotranspiration. By calculating the soil water balance of the root zone on a daily basis, the timing and depth of irrigation can be planned. To avoid crop water stress, irrigation water should be applied before or at the moment when the readily available soil water is depleted ($D_{mo,j} \geq RAW_j$). $D_{mo,j}$ is given by equation (2.12) and $RAW_j$ is given by equation (2.11). To avoid deep percolation losses that may leach relevant nutrients out of the root zone, the net irrigation depth should be smaller than or equal to the root zone depletion ($I_j \leq D_{mo,j}$). The daily net volume of irrigation is determined with the assumption that the crop fully evapotranspires without suffering from water stress throughout the day; the water volume is given by the following relationship,

$$I_j = D_{mo,j} - RAW_j + k_{c,j} \cdot ET_0 \qquad [mm]. \tag{2.15}$$

Irrigation increase the soil water content up to the minimum readily available water content (i.e., $D_{mo,j} = RAW_j$, see Figure 2.1c) that supplies the crop with the water required to satisfy its evapotranspiration demand throughout the day (namely $k_{c,j} \cdot ET_0$). In the evening, the root zone depletion is given by

$$D_{ev,j} = D_{mo,j} + ET_{a,j} - I_j \qquad \left[\frac{\text{mm}}{\text{day}}\right], \tag{2.16}$$

where $ET_{a,j}$ is given by equation (2.5) with $k_{s,j} = 1$. $ET_{a,j}$ is the water volume evapotranspired by the crop during the day; the water volume consists of green and/or blue water. The blue water, $ET_{b,j}$, corresponds to the irrigation water given to the crop (namely, $ET_{b,j} = I_j$); the green water, $ET_{g,j}$, is evaluated as the difference between the total and the blue water evapotranspiration.

### 2.1.3 Crop actual yield

Sub-national datasets of crop yields at high spatial resolution are seldom available. *Monfreda et al.* (2008) [94] and the FAOSTAT database provide good estimates of yield values. The first one refers to year 2000 providing the observed yields and harvested areas of 175 distinct crops on a 5 by 5 arc minute grid. This dataset has been widely used both for the $CWF$ assessment (e.g., [126]; [62]) and in analyses on crop yield-gaps (e.g., [86]; [95]). The FAOSTAT database provides annual yields at the country-scale from year 1961 to 2013. In order to obtain high-resolution actual yields, $Y_{a,T}$, referred to the investigated period, $T = [1996, 2005]$, we use the above mentioned data-sources combined in the following relationship as

$$Y_{a,T}(i,T) = g(i,2000) \cdot f(c,T), \tag{2.17}$$

where $g(i,2000)$ defines the ratio of cell yield to country yield for year 2000 ($i$ is the cell of the grid), and $f(c,T)$ describes the country-scale ($c$ indicates the country) yield in the investigated interval $T$. More specifically, $g(i,2000)$ is defined in each cell as

$$g(i,2000) = \frac{Y_a(i,2000)}{Y_a(c,2000)} \qquad [-], \tag{2.18}$$

where $Y_a(i,2000)$ is the yield measured in the cell in year 2000 (given by [94]), and $Y_a(c,2000)$ is the country-based yield in the cell for the same year (given by FAOSTAT). The function $f(c,T)$ is the average of the national yields given by FAOSTAT for each year $t$ of the study period, $T$, namely $f(c,T) = 1/10 \cdot \sum_{t=1996}^{t=2005} Y_a(c,t)$. In Figure 2.2, we provide the wheat yield as an example of the gridded yield used in this thesis.

Finally, the green and blue $CWF$ in each grid cell are determined with equation (2.1), substituting $ET_{a,y}$ with $ET_{g,y}$ for the green component and with $ET_{b,y}$ for the blue component $CWF$. The total virtual water content of the cell is the sum of the green and blue content (equation 2.3).

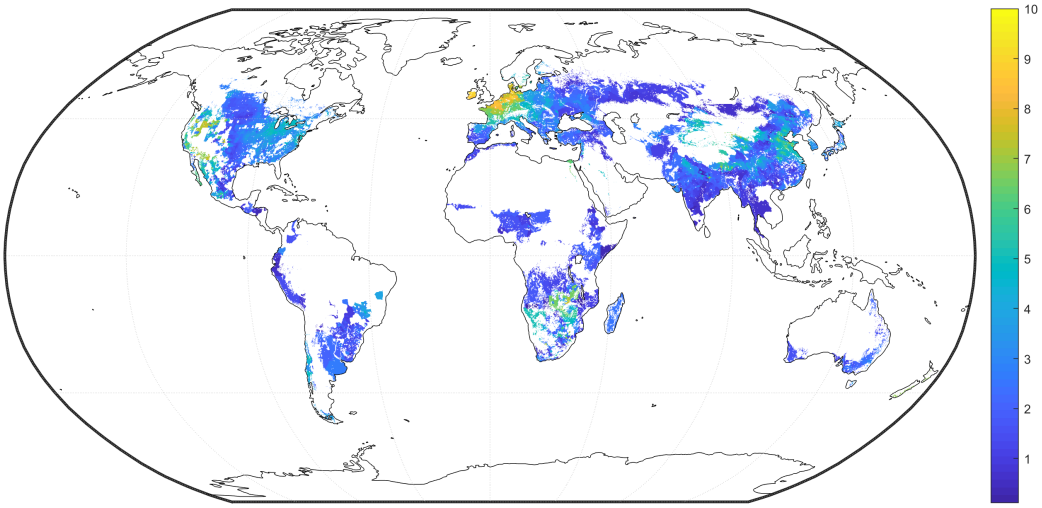


Fig. 2.2 Spatial distribution of the wheat yield [ton/ha] in the period 1996-2005.

### 2.1.4 Comparison with previous studies

The distributed results are aggregated at the country and global scale by a weighted mean (using cell production as weight) in order to make a comparison with the results from earlier studies. At national scale, our $CWF$ estimates are in good agreement with those from [62], particularly for maize production, as can be seen in Table 4.1 for the major exporting countries. National $CWF$ values estimated in the present work are also in close agreement with those from [89], as confirmed by the coefficients of determination, $R_w^2$, that are 0.91 for wheat, 0.76 for rice, 0.90 for maize, and 0.91 for soybean ($R_w^2$ is weighted with the countries production). At global scale, $CWF$ averages estimated in this study well compare with those from [126] and [89], especially for wheat and maize production, as can be seen in Table 2.2.

## 2.2 Sensitivity of crop water use to variability of input variables

A sensitivity analysis is required to understand how data uncertainty propagates through the virtual water content estimates and to identify the model-inputs that significantly affect the model-outputs. To this aim, a first-order sensitivity

Table 2.1 Comparison between the crop water footprint values ($CWF$) of wheat, rice, maize, and soybean, evaluated in the major exporting-countries, from the current study and from [62].

| Crop | Country | Current study [$m^3 \cdot ton^{-1}$] | [62] [$m^3 \cdot ton^{-1}$] |
|---|---|---|---|
| Wheat | USA | 2579 | 1359 |
| | Canada | 1275 | 1800 |
| | France | 668 | 366 |
| | Australia | 1877 | 1339 |
| | Argentina | 2034 | 1190 |
| Rice | Thailand | 2292 | 1831 |
| | Vietnam | 1675 | 1245 |
| | China | 1087 | 789 |
| Maize | USA | 657 | 621 |
| | Argentina | 918 | 1041 |
| | China | 725 | 715 |
| Soybean | USA | 2318 | 1921 |
| | Brazil | 2125 | 2220 |
| | Argentina | 1870 | 2405 |

Table 2.2 Comparison between the globally-averaged crop water footprint ($CWF$) of wheat, rice, maize, and soybean from the current study and from [89] and [126].

| Crop | Current study [$m^3 \cdot ton^{-1}$] | [89] [$m^3 \cdot ton^{-1}$] | [126] [$m^3 \cdot ton^{-1}$] |
|---|---|---|---|
| Wheat | 1523 | 1619 | 1469 |
| Rice | 1607 | 1486 | 1382 |
| Maize | 933 | 1028 | 1089 |
| Soybean | 2258 | 2107 | 2406 |

analysis is applied: the functional dependence of $CWF$ on each input parameter is expanded as a Taylor series and truncated at the first order; in this way a linear relationship between the $CWF$ estimate and the generic input parameter, $x$, is assumed in a small neighbourhood of $x$. Parameters are perturbed one-at-a-time of a very small quantity, which is arbitrarily chosen. To evaluate and compare the sensitivity of the $CWF$ to different parameters, we define a normalized sensitivity index, $SI_x$, for each parameter, $x$, as

$$SI_x = \left(\frac{\Delta CWF}{CWF_0}\right) \Big/ \left(\frac{\Delta x}{x_{ref}}\right), \tag{2.19}$$

where $\Delta CWF$ is the virtual water content variation resulting from changing the parameter $x$ of a quantity $\Delta x$. Both variations are normalized: $\Delta CWF$ with respect to the virtual water content ($CWF_0$) estimated when all parameters are at their baseline values, while $\Delta x$ with respect to a reference value of the parameter, $x_{ref}$ (see Section 2.2.1). Positive and negative variations of the input parameters are considered to analyse the response of the $CWF$ both in terms of magnitude and direction of the change. The sensitivity analysis focuses on four key input parameters, namely monthly reference evapotranspiration ($ET_{0,m}$), available water content ($AWC$), planting date ($PD$), and length of the growing period ($LGP$). The reference evapotranspiration and the available water content are varied by $\pm 0.01$ mm day$^{-1}$ and $\pm 1$ mm/m respectively, while the planting date and the length of the growing are changed by $\pm 1$ day. The imposed variations are different from parameter to parameter, depending on their standard deviation, average, and range of variation. All changes are lower than 2% of the standard deviation and lower than 1% of the range, in order to guarantee that variations are small.
For each single variation, the new $CWF$ value is calculated. The $CWF$ variation ($\Delta CWF$) is due to a variation of the water volume evapotranspired by the crop during the growing season ($ET_{a,LGP}$) and to a variation of the crop actual yield ($Y_a$). The new evapotranspiration, $ET'_{a,LGP}$, is determined by the equations (2.4,2.5) where the modified parameter is introduced. The new yield, $Y'_a$ (whose variation is affected by the evapotranspiration change), is determined through a modified expression of the equation by [38],

$$\left(1 - \frac{Y'_a}{Y_a}\right) = K_y \cdot \left(1 - \frac{ET'_{a,y}}{ET_{a,y}}\right), \tag{2.20}$$

where $K_y$ is the yield response factor, representing the effect of a reduction in evapotranspiration on yield losses, $Y_a$ and $Y'_a$ are the actual yields before and after the input parameter change, respectively, and $ET_{a,y}$ and $ET'_{a,y}$ are the crop actual evapotranspiration before and after the change of the input parameter.

### 2.2.1 The normalization of input parameters

The normalization of the parameter variation required by the sensitivity index defined in equation (2.19) is different for the four parameters analysed. In detail, the variation of the available water content ($AWC$) is normalized with respect to the baseline value; the variation of the planting date ($PD$) is normalized with respect to 360 days (namely the number of days of a year considering each month 30 days long); the variation of the reference evapotranspiration ($ET_{0,m}$) and the length of the growing period ($LGP$) are normalized as follow. After changing their baseline values of a fixed quantity (0.01 mm day $^{-1}$ and 1 day, respectively), the new virtual water content of the rainfed and the irrigated production, as well as the relative variations $\Delta CWF$, are evaluated. The normalized sensitivity index are separately determined for rainfed and irrigated conditions, with specific values of $x_{ref}$, that are (i) the ratio between the total reference evapotranspiration over the growing period and $LGP$ for $ET_{0,m}$, and (ii) the length of the growing season typical of rainfed and irrigated conditions for $LGP$. Finally, the overall sensitivity indexes of these parameters, $SI_{ET_{0,m}}$ and $SI_{LGP}$, are calculated as the weighted mean of the rainfed and irrigated sensitivity indexes, using the harvested area given by [108] as the weights.

## 2.3 Results of the water footprint assessment

The water footprint assessment focuses on four widely cultivated crops (wheat, rice, maize, soybean) which provide around 50% of the global caloric content and 47% of the proteins in global human food consumption [34]. Annual crop production, averaged over the period 1996-2005, requires 2640 $km^3 \cdot yr^{-1}$ of water, with 82% of water coming from soil moisture and 18% from surface and ground.

### 2.3.1 The green and blue water footprint of wheat

Figure 2.3 reports the spatial distribution of the total and blue crop water footprint of wheat, typical of the period between 1996 and 2005. The maps show a strong spatial heterogeneity both inside the climatic regions and at the sub-national scale. The observed spatial variability is mainly driven by the

yield pattern with a correlation coefficient of 0.74, while the influence of the evapotranspiration demand (hence, climate), is lower with a correlation of 0.34. Looking at the maps, one can immediately notice the high water efficiency (low $CWF$) of the United States (especially on the West Coast), Europe, and China, where the virtual water content is generally lower than 2000 $m^3$ $ton^{-1}$, with a consumption of blue water less than 10% of the total water consumption in the United States and Europe, and between 50% and 75% in the large cropping area in the north-east of China. South America, Africa, and Southern Asia are less water efficient, with $CWF$ reaching up to 6000-8000 $m^3$ $ton^{-1}$ in some regions of Venezuela, Ethiopia, and Vietnam.

**Asia.** China and India are the largest producers of wheat in the study

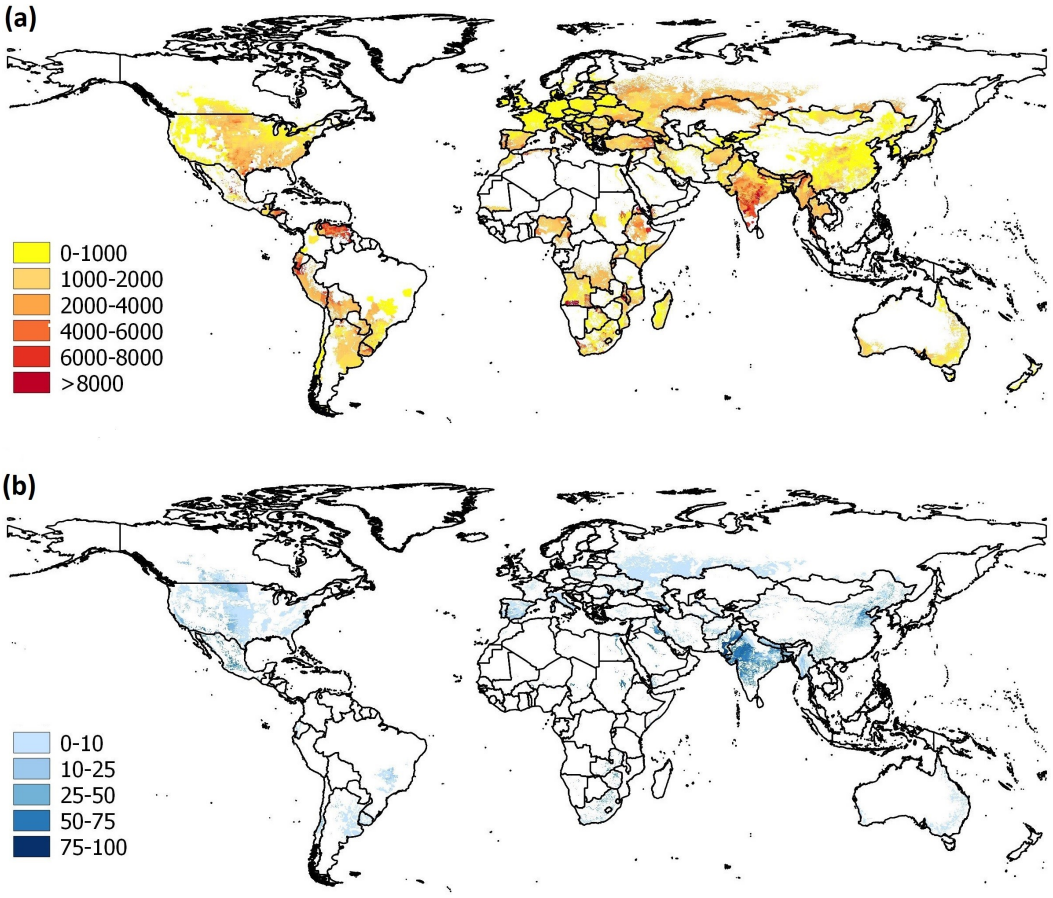


Fig. 2.3 Spatial distribution of the crop water footprint ($CWF$) of wheat in the period 1996-2005: (a) total $CWF$, expressed in $m^3$ $ton^{-1}$ and (b) blue $CWF$, expressed as percentage of the total $CWF$.

period, accounting for 21% and 12% of the global production, respectively.

Wheat is grown throughout China with virtual contents range between 500 and 6000 $m^3$ $ton^{-1}$, but production is mostly located in the North-East of the country, where $CWF$s are around 1000 $m^3$ $ton^{-1}$, thanks to the higher yields achieved (around 6 ton $ha^{-1}$) and lower evapotranspiration demand than the rest of the area. This region falls, in fact, in the temperate belt where yearly $ET_0$ is approximately 1000 mm [98]). In this region more than 75% of the total $CWF$ is blue (2.3(b)), meaning that water from precipitation gives only a small contribution to meet crop water requirement, while irrigation plays a strategic role in determining high crop water productivity (e.g., tons of wheat produced per volume of water used). Water productivity decreases in Central and Southern China where $CWF$s are within 2000-6000 $m^3$ $ton^{-1}$. Indian wheat production is more water consuming than Chinese one. The different climatic conditions and the local lower yields, in fact, make India less water efficient. Even though wheat is grown almost everywhere in India, about 90% of the total production lies in the Ganges and Indus basins, which are the wheat baskets of India and Pakistan. Both the basins have a higher water productivity, as shown by the low $CWF$ (1800 $m^3$ $ton^{-1}$ on average), than other wheat producing areas (2.3 1(a)), thanks to the higher yields and lower evapotranspiration demands. Moreover these lands are mostly irrigated, allowing the crop to fully evapotranspire with blue water generally higher than 50% of the total water content. The rest of Asia is characterized by $CWF$s in the range of 2000-6000 $m^3$ $ton^{-1}$ or even higher in the tropical zone, where yields are lower. A possible explanation of such low yields can be found in the wheat-rice double cropping. Since wheat tends to be less profitable than rice, farmers may prefer to achieve an optimal harvesting date for rice, planting wheat later than the optimal time [29]. This is the case of Pakistan, for example, where about 50% of wheat production is grown after rice or cotton.

**America.** The United States is the third main wheat producer (11% of the global production). Production is mostly located in the North American plains –the Wheat Belt– where the $CWF$ shows the highest spatial variability of the region. Yields are around 2 ton $ha^{-1}$, and crop evapotranspirations range from 300 mm in the North to 700 mm in the South, following the climate pattern (correlation coefficient of 0.44).

The cropping areas extends throughout in the USA, showing $CWF$s of 1000-4000 $m^3$ $ton^{-1}$ in the East Coast (where a net separation between the temperate

and sub-tropics zone can be notice) and 500-2000 $m^3$ $ton^{-1}$ in the West Coast. The sharp differences in coastal virtual water content reflect the differences in yields –around 3.5 ton $ha^{-1}$ in the East and over 5 ton $ha^{-1}$ in the West– and evapotranspiration pattern –that shows an average value of 750 mm in the East and 450 mm in the West. The West Coast registers lower $ET$ values because wheat is planted later than in the East Coast and the growing period is shorter. In Central America the only countries with significant production are Mexico and Guatemala [29]. In Mexico, all wheat is produced under irrigation, as shown by the high percentage of blue water contents in 2.3(b), reaching 80% in many cells. In Guatemala, wheat is mostly produced in highland valleys under rainfed conditions. Here, diseases adversely affect the crop under the usually wet production conditions, and weeds can significantly constrain production [29], as a consequence yield is as low as 1 ton $ha^{-1}$ and $CWF$ reaches 8000 $m^3$ $ton^{-1}$ in some cells. In South America the spatial variability of the $CWF$ is remarkable, and mostly driven by yield (correlation of 0.90): from 2000 $m^3$ $ton^{-1}$ in Brazil and Argentina ($Y_a = 3$ ton $ha^{-1}$ on average) to 8000 $m^3$ $ton^{-1}$ in Venezuela and Bolivia ($Y_a < 0.5$ ton $ha^{-1}$), with intermediate values in Peru and Colombia. The generally low values are due to many production constraints, such as moisture availability, diseases, poor cultural practices, and lack of fertilizers, pesticides and farm implements [29]. Only few Brazilian cells show virtual water contents around 1000 $m^3$ $ton^{-1}$ thanks to the improved cultural practices and disease resistance.

**Europe.** European lands present very low $CWF$ values with only few exceptions. The temperate-oceanic zone (i.e., Northern and Western Europe) shows values lower than 1000 $m^3$ $ton^{-1}$ (e.g., France, Germany, and Northern Italy), or even lower than 500 $m^3$ $ton^{-1}$ (e.g., United Kingdom, Sweden, and Norway). These lands present yields around 7-8 ton $ha^{-1}$ (the highest in the world) and total evapotranspiration of about 500 mm. The spatial distribution of $CWF$ is one of the most homogeneous in the world since the local $ET_0$ variations are small and the yields are mostly the same in all countries. In Eastern and Southern Europe values are generally around 2000 $m^3$ $ton^{-1}$, with the exception of Portugal where $CWF$ reaches 4000 $m^3$ $ton^{-1}$, being the yield lower than 1 ton $ha^{-1}$. Generally, in Europe the virtual water content of wheat is mostly green (blue $CWF$ is always lower than 10% of the total content). It follows that Europe is not only very efficient in term of water productivity (it has

the lowest $CWF$ of the world), but it is able to produce wheat with a limited depletion of blue water.

**Oceania.** Australian tropics land is only harvested in the Eastern coast with virtual water contents around 1000 $m^3$ $ton^{-1}$, thanks to low $ET$ and average yield of 3 ton $ha^{-1}$. Most wheat is grown in the arcuate belt of land curving across the eastern and southern regions where winter rainfall generally satisfy crop water requirement (blue $CWF$ is only 5% of the total one, as shown in 2.3(b)).

Wheat is a major crop in New Zealand with very low $CWF$ thanks to the high yields: 6 ton $ha^{-1}$ on a country average.

**Africa.** The African continent is really heterogeneous, with virtual water contents ranging from 1000 to 8000 $m^3$ $ton^{-1}$. Algeria, Egypt, Ethiopia, Kenya, Morocco, South Africa, and Tunisia are the main producers, with production around $5{\cdot}10^5$ tonnes in 2010 [50]. These countries exhibit very different $CWF$s, determined by significantly different yields and reference evapotranspiration (correlation of 0.75 for yield and 0.40 for evapotranspiration). Algeria and Morocco show different virtual water contents –3350 and 1500 $m^3$ $ton^{-1}$ respectively– mainly because of the different yields. Both countries, in fact, have increased their productions of about 15% during the last decade [50], but while the performance of Morocco was due to improving yields up to 5 ton $ha^{-1}$, resulting in lower $VWC$s, Algeria's larger production was due to increasing harvested areas with a constant average yield of 2.5 ton $ha^{-1}$ and higher $VWC$s.

Kenya and Ethiopia have similar yields, 2 and 3 ton $ha^{-1}$ respectively, but really different $VWC$s, 1500 $m^3$ $ton^{-1}$ for Kenya (100% green) and 4200 $m^3$ $ton^{-1}$ for Ethiopia (100% green). The reason lies in the crop actual evapotranspiration distribution that shows homogeneous values of about 700 mm in Ethiopia and in the range of 200-500 mm in Kenya. Ethiopian growing season is, in fact, 50 days longer than Kenyan one. Among the main producers, South Africa is the most heterogeneous country with values from 1000 to 8000 $m^3$ $ton^{-1}$. Both yield and evapotranspiration have remarkably high spatial variability, from 0.5 to 6 ton $ha^{-1}$ for $Y_a$, and from 200 to 700 mm for $ET$. A possible explanation is the different irrigation conditions of the harvested area, shown in 2.3(b). Finally, Egypt shows the smallest $VWC$ of all African lands, 1100 $m^3$ $ton^{-1}$, with the blue component accounting for 100% of the total $VWC$. Wheat is, in fact, grown only in the Nile basin where, thanks to the silt deposits from the

Nile, average yield is 7 ton $ha^{-1}$ (similar to the one achieved in North and West Europe).

### 2.3.2 The green and blue water footprint of rice, maize, and soybean

**Rice.** The spatial distribution of the virtual water content of rice is reported in Figure 2.4. More than 90% of the world's rice comes from Asia, with China and India as the lead producers (42% and 21% of the global production). Similarly to wheat, China is much more water efficient than India: i.e., 904 $m^3$ $ton^{-1}$ (40% blue) and 1894 $m^3$ $ton^{-1}$ (38% blue). Both countries show a significant blue virtual water content that together with the hybrid species used makes rice better performing than wheat. In the central and southern Chinese wheat-growing areas, where more moisture is available, rice-wheat rotation is common [29]. Other producing areas are located in the Eastern and Southern Asia (e.g., Indonesia, Bangladesh, and Vietnam), where virtual water contents are around 2000 $m^3$ $ton^{-1}$, or higher. In America, rice production is limited to the South of the continent (there are only few harvested areas in Northern and Central America, e.g. i the Mississippi river basin), where all 13 countries grow rice. The average yield is approximately 5 ton $ha^{-1}$ [51], but there is a large variability between and within the countries, that makes the spatial distribution of $CWF$ really heterogeneous (correlations coefficient of 0.94). The yield gap observed is the result of numerous deficiencies, as explained for wheat. Similarly to wheat, European rice production shows low and homogeneous virtual water contents, which are generally lower than 1000 $m^3$ $ton^{-1}$, with blue water playing a significant role, as shown in Figure 2.4(b). With the exception of few cells located in Morocco, Algeria, Egypt, and Madagascar, Africa exhibits high virtual water contents. Insufficient technologies, ineffective farmers organization, and environmental constraints are the main factors limiting rice yields. In particular, weed is the most important biotic factor reducing rice production in Sub-Saharan Africa, followed by rice blast disease caused by a fungus [119]. Moreover, the share of irrigated rice area in Africa is very small, resulting in a significant water stress during the growing season. Egypt is again an exception with 100% of the cultivated area irrigated. Australia is harvested only in a small region in the South East, with virtual water content lower

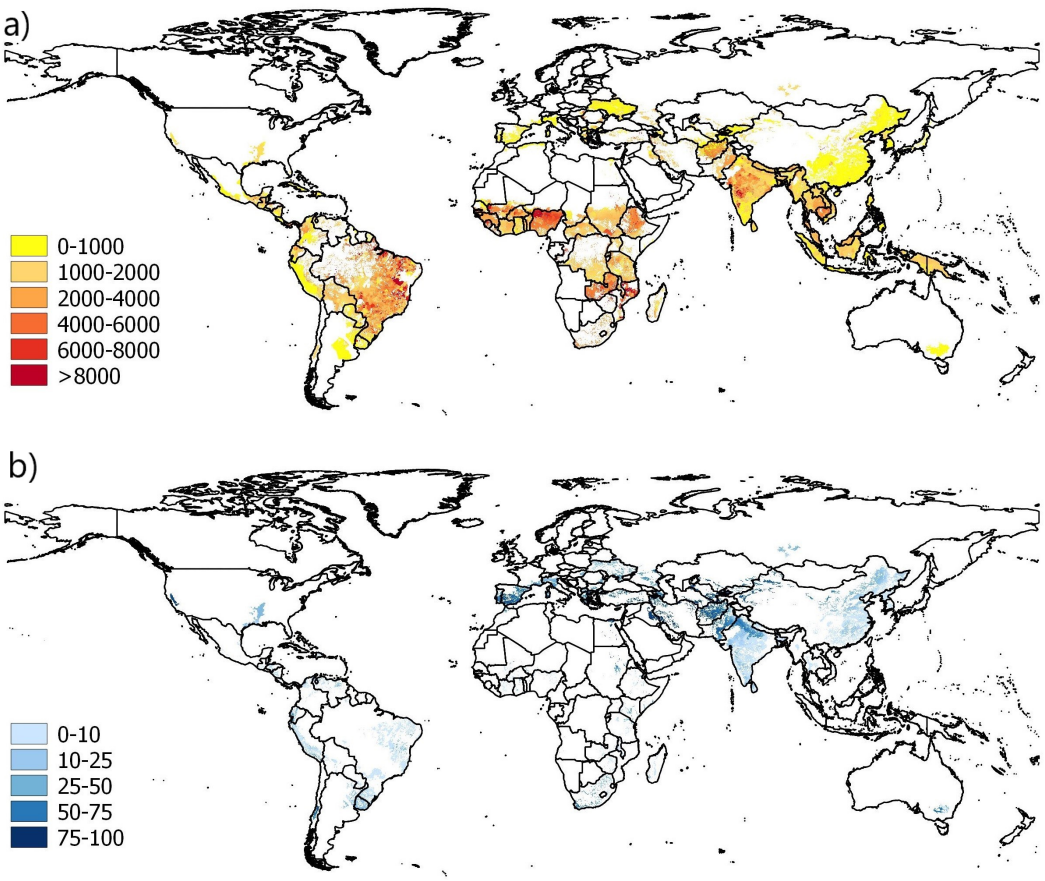


Fig. 2.4 Spatial distribution of the crop water footprint ($CWF$) of rice in the period 1996-2005. Total $CWF$, expressed in $m^3$ $ton^{-1}$ (a); Blue $CWF$, expressed as percentage of the total$CWF$ (b).

than 1000 $m^3$ $ton^{-1}$. Australia is the region of the world with the highest yield (9 ton $ha^{-1}$), thanks to the high yielding rice varieties that use less water and to the excellent water management practices. According to the Government (Department of Agriculture), the Australian rice industries lead the world in water use efficiency since they use 50% less water than the global average.

**Maize.** Maize production is much more widespread worldwide; in fact, it can be cultivated in extremely cool, moderate, and very hot climate, under very different moisture regimes [123]. Along the temperate belt, maize has the highest water use efficiency as shown in Figure 2.5(a) by the $CWF$<1000 $m^3$ $ton^{-1}$; conversely, across the tropics belt the virtual water content are larger due to lower yields. In these areas, maize goes under water stress conditions more often due to the absence of irrigation, which feeds maize less than 10% of the

water requirement. The main production is given by the United States (40% of the global production), that is characterized by virtual water contents around 645 $m^3$ $ton^{-1}$ or even smaller, followed by China (18%) where the average $CWF$ is 767 $m^3$ $ton^{-1}$. Brazil is the third largest producer (5% of the global production) in the world, with average $CWF$ of 1670 $m^3$ $ton^{-1}$. Differently from

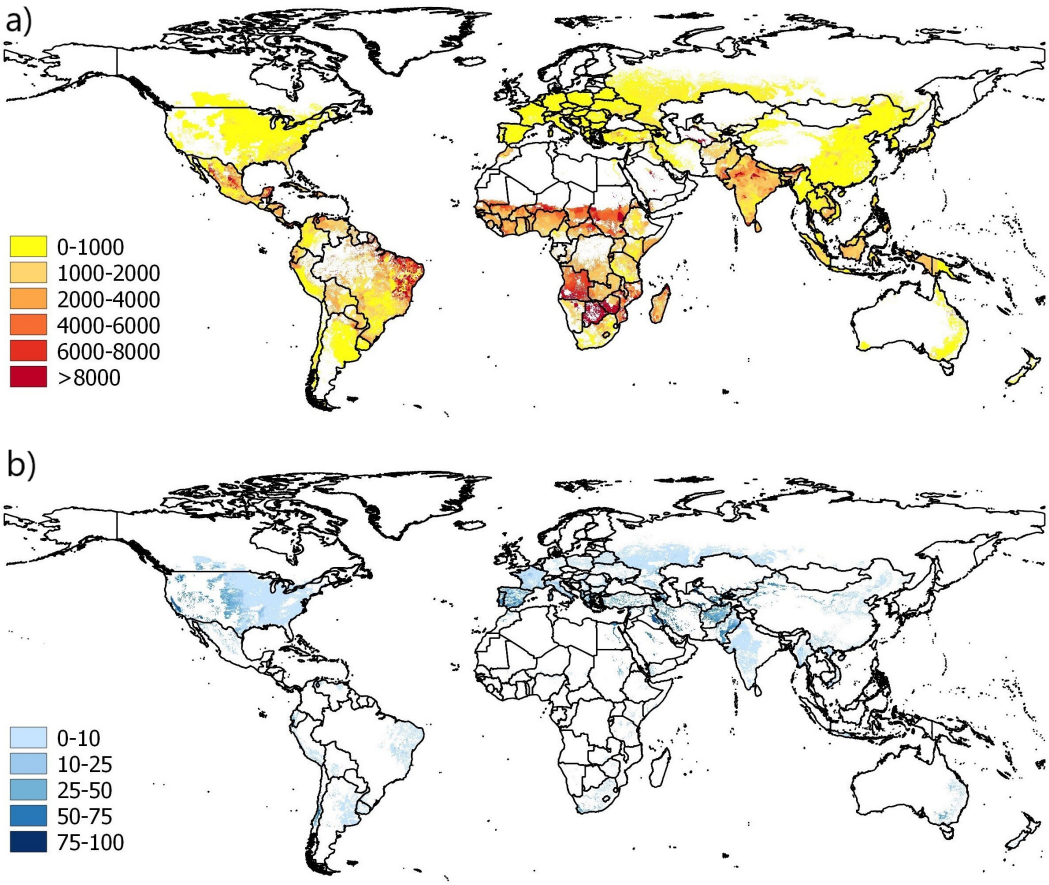


Fig. 2.5 Spatial distribution of the crop water footprint ($CWF$) of maize in the period 1996-2005. Total $CWF$, expressed in $m^3$ $ton^{-1}$ (a); Blue $CWF$, expressed as percentage of the total $CWF$ (b).

the US and China, Brazil shows larger spatial variability of water use efficiency: $CWF$s are higher along the east coast, and lower in the South. Mexico gives also a significant contribution to the global production, with an average $CWF$ of 1312 $m^3$ $ton^{-1}$) and yield values smaller than 3.5 ton $ha^{-1}$. Also for maize, Asia results really heterogeneous, but, differently from the case of wheat, it shows a smaller range of variability. Maize is mainly grown in China, where virtual content, as well as yields, are homogeneously distributed.

France is the biggest European producer (25% of the European production in year 2000, according to [51]) with an average $CWF$ lower than 500 $m^3$ $ton^{-1}$ (the blue component is 12% of the total content). All Europe is characterized by values lower than 1000 $m^3$ $ton^{-1}$. Australia grows maize in the same lands of wheat, but maize achieves bigger yields, resulting in slightly lower virtual water content. South Africa ($CWF$ of 1606 $m^3$ $ton^{-1}$, 1% blue) is the biggest African producer, with the second biggest yield after Egypt ($CWF$ of 860 $m^3$ $ton^{-1}$, 87% blue) that contributes to half the South African production. The rest of Africa mostly presents $CWF$ in the range of 4000-6000 $m^3$ $ton^{-1}$.

**Soybean.** From the map in Figure 2.6(a), it is clear that soybean is typically characterized by higher water contents than other crops. Abiotic and biotic stresses, in fact, may limit soybean production worldwide. The United States is the main producer (47% of the global production), with cultivated areas mostly located in the north plains and on the East Coast, where virtual water content reaches 4000 $m^3$ $ton^{-1}$. Soybean is mostly reliant upon green water, but close to the High Plain aquifer around 25-50% of the water content is contributed by the groundwater (see Figure 2.6(b)). The yield is almost the same in all the American continent (around 2-3 ton $ha^{-1}$) and the highest values are observed in Brazil and Venezuela. In particular, Venezuela presents a lower $CWF$ than Brazil because it is characterized by a shorter growing period and thus a lower evapotranspiration demand. American virtual water content distribution is correlated to yield and evapotranspiration with the same correlation coefficient of 0.70. Europe shows virtual water contents similar to South America, or even lower in countries like France and Italy where larger yields are achieved and where irrigation accounts for 25% of the crop water use. Asia appears largely heterogeneous with $CWF$s ranging from 1000 to 8000 $m^3$ $ton^{-1}$. China and India are again the leading producing countries of Asia, but, differently from the other crops, their water efficiency is lower. This is particularly evident for China where the national average $CWF$ is about 2500 $m^3$ $ton^{-1}$, which is higher than the global average. On the other hand Kazakhstan, Turkey, and Iran show lower virtual water contents thanks to the higher yields. Virtual water contents are much higher in Africa where peaks over 8000 $m^3$ $ton^{-1}$ are found in Nigeria (primary African producer of soybean) and Tanzania.

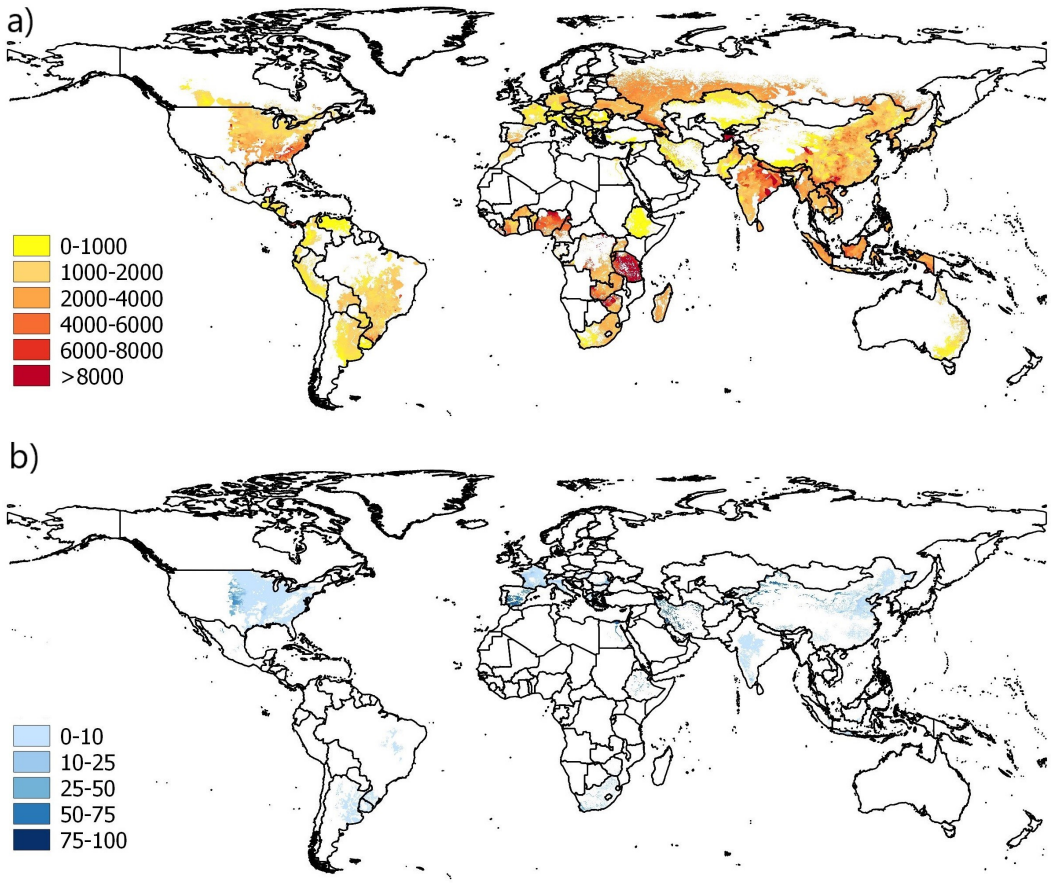


Fig. 2.6 Spatial distribution of the crop water footprint ($CWF$) of soybean in the period 1996-2005. Total $CWF$, expressed in $m^3$ $ton^{-1}$ (a); Blue $CWF$, expressed as percentage of the total $CWF$ (b).

### 2.3.3 Distribution of $CWF$ related to the production of wheat, rice, maize, and soybean

We evaluated the distribution of $CWF$ as a function of yearly crop production typical of the study period (Figure 2.7), where crop production is given by the multiplication of crop actual yield and harvested area. At the global scale, the histograms of wheat and rice are skewed towards the right with tailing off after 5000 $m^3$ $ton^{-1}$. Both crops show a high water productivity, but wheat has a larger water consumption since its production is bigger and more widespread worldwide (the average global water consumption in the considered period is about 900 $Gm^3$ $yr^{-1}$, 86% of which is green). Rice presents a lower water consumption of 870 $Gm^3$ $yr^{-1}$ (64% is green). Maize exhibits a skewed

right pattern with tailing off after 4000 $m^3$ $ton^{-1}$. Its production, which is the biggest one among all of these four crops, is the most efficient in terms of water consumption (830 $Gm^3$ $yr^{-1}$, 95% is green). Soybean is the most water consuming crop per tons of product; however, since it is less produced than other crops, it contributes to the smallest total water consumption (400 $Gm^3$ $yr^{-1}$, 97% is green). Figure 2.7(a) shows that Asia is the main wheat producer ($3 \cdot 10^8$ ton $yr^{-1}$), and exhibits the highest virtual water variability. Europe and North America are also important contributors to the global wheat production (46% and 27% of Asian production), with a smaller $CWF$ range. Asia and Europe are the most efficient regions in terms of water consumption since their production is mostly characterized by low $CWF$ values: the histograms are, in fact, skewed towards the right. However, while Europe uses nearly no blue water for wheat production (99.5% is green water), in Asia about 20% of the water footprint of wheat is contributed by blue water. North America appears to be less water efficient, because the core of its production has a higher water footprint; for example, it produces $6 \cdot 10^7$ ton/y less than Europe using 3% more water. Finally, Africa, Oceania, and South America are minor wheat producers (around 7% of Asian production).

In the case of rice (Figure 2.7(b)), Asia is not only the main producer, but its production is significantly larger than that from all the other geographic areas (it accounts for 93% of the global production). Asian rice production is rather water efficient: its histogram pattern is skewed towards the right, with virtual water content mostly lower than 2000 $m^3$ $ton^{-1}$. This high efficiency is mostly due to China (which is the biggest producer), where high yields –6.5 ton $ha^{-1}$ on average– are achieved by many varieties and hybrids with good quality and resistance to diseases and insects [84]. Among the other geographical areas, only America and Africa give an appreciable additional contribution to world production (5.2% and 2.4% of Asian production). In particular, 40% of North America water consumption for rice production is blue, while in South America and Africa blue water contributes to the water footprint of rice only for 10% and 13%, respectively, and have an overall lower water efficiency. In relative terms, in Oceania the blue water footprint of rice is higher than in all the other continents, with more than 70% of the water consumed coming from irrigation.

Ninety percent of the global production of maize is located in Asia and America (Figure 2.7(c)), with the United States leading the water efficiency

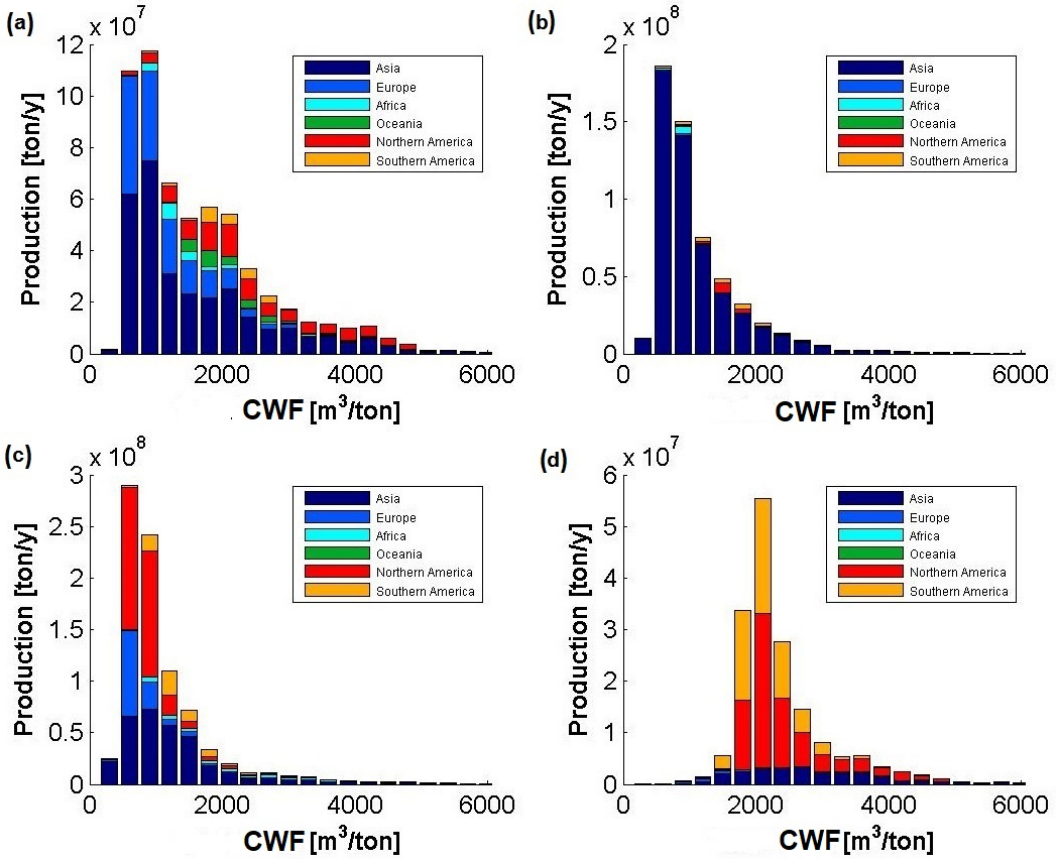


Fig. 2.7 Distribution of the total crop water footprint ($CWF$) related to yearly production: (a) wheat; (b) rice; (c) maize and (d) soybean.
In each histogram the abscissa reports the $CWF$ grouped in classes of 300 $m^3$ $ton^{-1}$ width; the height of the rectangle gives the yearly production typical of the period 1996-2005 (i.e., crop actual yield multiplied by harvested area) for each class and geographical area, rectangle area indicates the volume of water used. We separate the contribution of North and South America in correspondence of Panama.

of the region with an average water footprint of 645 $m^3$ $ton^{-1}$. The $CWF$ of North America shows a small spatial variability, with values generally lower than 1000 $m^3$ $ton^{-1}$ and mostly contributed by green water. Conversely, Asia is characterized by a higher heterogeneity similarly to the one of wheat and rice. Europe (where 15% of the global production is located) exhibits an efficiency similar to the one observed in North America both in terms of total consumed water and in terms of green contribution. The Americas are the biggest producers of soybean (Figure 2.7(d)). North and South America present similar $CWF$ distributions, with a higher variability in North America. Asia is also an important producer, with an overall symmetric distribution of virtual water contents, indicating a lower water efficiency compared to other crops.

### 2.3.4 Statistical distribution of CWF related to the production of wheat, rice, maize, and soybean

Box-plots in Figure 2.8 compare the average $CWF$ of different crops, grouping data at a continental and global scale, and highlight the associated variability as a function of production. The $CWF$ values calculated at the pixel level (see equation 2.1) within each continent are sorted in an ascending-order vector which is then used to sort the cumulated percentage of cell production values. Quartiles are determined in correspondence to 25%-50%-75% of the cumulated production. The continental and global averages are obtained as a production weighted mean of the pixel values (i.e., cell production is used as the weight in the average). At a global scale, maize is the crop with the lowest $CWF$, 927 $m^3$ $ton^{-1}$ on average. Except for the case of Africa, all geographic areas show a $CWF$ lower than 1500 $m^3$ $ton^{-1}$ for, at least, 75% of total production, or even lower than 800 $m^3$ $ton^{-1}$ in Europe, Oceania, and North America. Moreover, most of the areas exhibit a relatively low spatial variability (with Europe and North America being the most homogeneous regions). On the contrary, African virtual water contents range from 1000 (e.g., Egypt, Kenya, Madagascar, and South Africa) to over 3000 $m^3$ $ton^{-1}$ (e.g., Ethiopia, Nigeria, and Congo). The lowest $CWF$s are found in those regions where irrigation prevents the crops from water stress and yield reduction, or in those areas where high-yielding genetically modified maize is planted, as in South Africa [123].

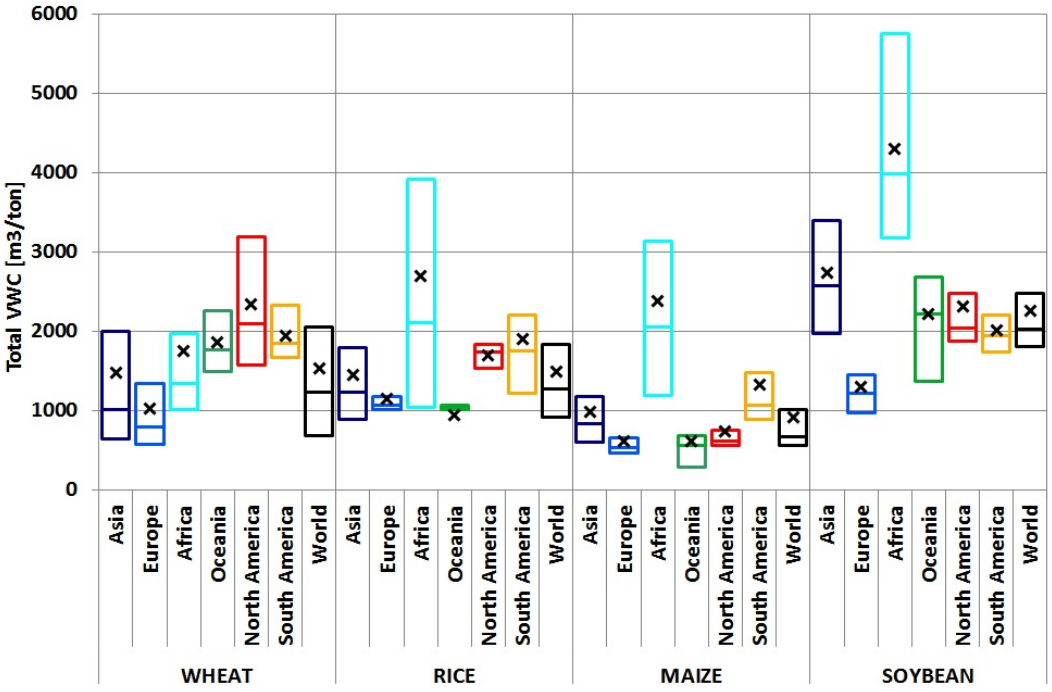


Fig. 2.8 Boxplots and weighted means (represented by cross markers) of the virtual water content of wheat, rice, maize, and soybean aggregated by continents. Boxplots have been obtained considering virtual water content cell values in ascending order together with the relative cumulative production and quantiles have been determined in correspondence to 25%-50%-75% of total production.

Rice shows an average virtual water content of 1435 $m^3$ $ton^{-1}$. As expected, its global $CWF$ range is very similar to the one found for Asia (which is the biggest producer). Similarly to maize, Africa presents the largest virtual water content (2700 $m^3$ $ton^{-1}$) and the widest interquartile range (1000-4000 $m^3$ $ton^{-1}$). Europe, Oceania and North America exhibit a very low spatial variability, which is explained by their homogeneous harvested areas where little differences in yields and evapotranspiration demands can be observed.

Wheat shows an average virtual water content of 1529 $m^3$ $ton^{-1}$. In this case, all the geographical areas show a quite similar $CWF$ spatial variability (as shown by the similar interquartile ranges in Figure 2.8): wheat is, in fact, the most widely cultivated cereal in the world with more than $220 \cdot 10^6$ ha planted annually [124]. Thus, wheat is grown under a wide range of climatic conditions (i.e., evapotranspiration patterns), soil properties, and production methods (i.e., yield patterns), which determine the wide range of $CWF$s. North America shows the highest $CWF$. African water efficiency is mostly determined by Egypt (i.e., leading the African production). Thanks to the fertility of the Nile Valley, Egypt can achieve wheat yields similar to those of Europe, the region where wheat production has the highest water efficiency.

Finally, soybean exhibits the highest virtual water content, 2243 $m^3$ $ton^{-1}$ on global average. Such a high value is mostly due to North and South America and, to a lesser extent, Asia. All the geographical areas have an average $CWF$ shifted above 2000 $m^3$ $ton^{-1}$, except for Europe that appears to be the most water efficient region, similarly to the cases of wheat and maize. Conversely, African values are really high, with an average $CWF$ value of 4250 $m^3$ $ton^{-1}$.

## 2.4 Results of the sensitivity analyses

The sensitivity analysis gives important insights into the model performance in terms of key input parameters. Positive and negative variations of the input parameters are found to produce $CWF$ variations of the same magnitude, but in opposite directions. Therefore, here we provide the sensitivity indexes, $SI_x$ for parameter $x$, only with respect to positive variations. In Figure 2.9, we report the average sensitivity indexes, $SI_x$, (evaluated with equation (2.19)) at the global scale, box-plots quantiles are referred to production with the same

approach used for the box-plots in Figure 2.8 (see Section 2.3.4). Figure 2.12 shows the sensitivity of wheat $CWF$ to the $LGP$ variation, $SI_{LGP}$.

### 2.4.1 Available water content

The available water content ($AWC$) is the difference between the water content at field capacity and wilting point. It varies across the grid cells with values from 3.75 to 150 mm $m^{-1}$, depending on the soil properties. This parameter, together with the rooting depth ($Zr_j$) and the depletion fraction ($\rho$), defines the readily available water content ($RAW$) that the crop can use to evapotranspire without experiencing water stress (see the equations (2.10,2.11)). Thus, $RAW$ is the water content defining incipient stomata closure and transpiration reduction ($k_s < 1$, see equation (2.9)). In the sensitivity analysis, the available water content is varied by $\pm 1$ mm $m^{-1}$. The variation in $AWC$ produces a variation of $RAW$, and thus a shift of the initial stomata closure to different water contents. For example, an increment of $AWC$ implies that crop becomes water stressed later during the growth season, resulting in a larger actual evapotranspiration ($ET_{a,LGP}$). In detail, green evapotranspiration increases while blue evapotranspiration decreases. The reduction of the blue component is explained by the lower irrigation requirement, since the condition $D_{mo,j} \leq RAW_j$ is satisfied for a longer period; as a consequence, the green component increases.

All crops (see Figure 2.9) exhibit negative sensitivity indexes ($SI_{AWC}$); a negative $SI_{AWC}$ means that an increment of the $AWC$ produces a reduction of the crop $VWC$ due to the higher achieved yields (see equation 3.5) through the higher evapotranspiration rates. The virtual water content of rice is the least sensitive to $AWC$ variations. Rice water stress is, in fact, controlled by irrigation (75% of the total harvested area is, in fact, equipped for irrigation), thus the positive effect of increasing the available water content is limited and the increase of $ET$ and $Y_a$ is merely appreciable. For wheat, maize, and soybean the water content is more substantially influenced by the $AWC$ and precipitation. In fact, these crops are less frequently irrigated (<30% of the total harvested area), and thus they are more influenced by the soil water conditions.

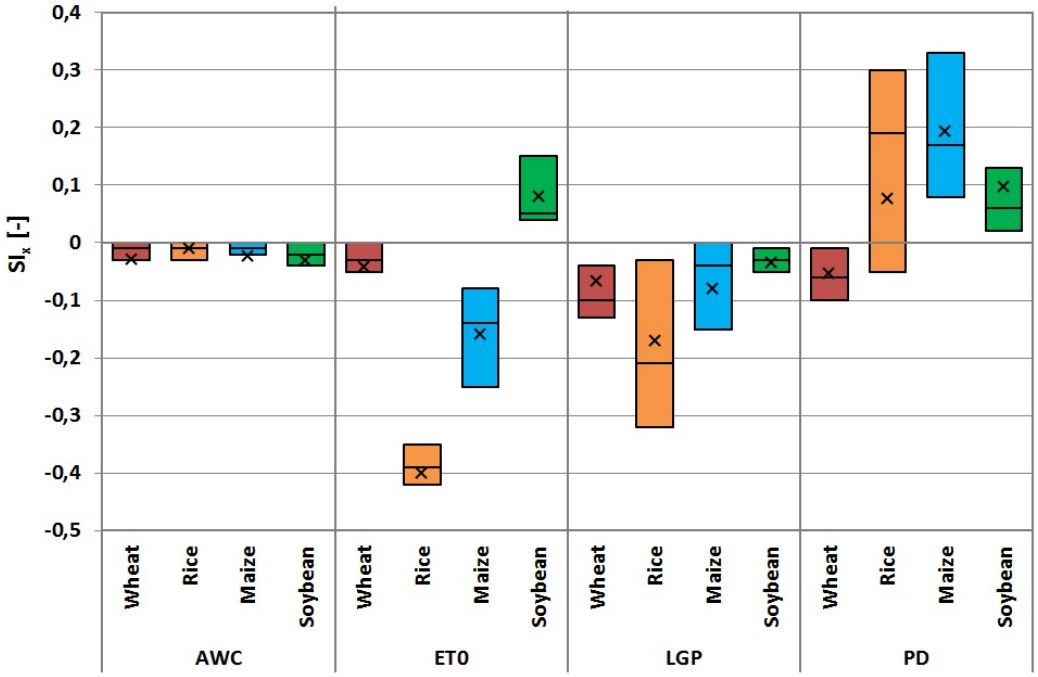


Fig. 2.9 Boxplots and average values (represented by cross markers) of the sensitivity index ($SI_x$) of each parameter $x$. $AWC$: available water content, $ET_{0,m}$: reference evapotranspiration, $LGP$: length of the growing period, $PD$: planting date.

### 2.4.2 Reference evapotranspiration

Monthly reference evapotranspiration, $ET_{0,m}$, is cell specific and represents the daily average evapotranspiration during a given month. The linear interpolation of $ET_{0,m}$ gives the daily temporal evolution of the reference evapotranspiration, $ET_{0,j}$, during the year, with $j$ being the day of the year. Planting and harvesting date, which are crop and cell-specific, define the range of $j$. The $ET_{0,j}$ value, multiplied by the daily crop coefficient ($k_{c,j}$), determines the daily evapotranspiration demand, which directly influences the crop virtual water content.

In the sensitivity analysis, $ET_{0,m}$ is varied by $\pm$ 0.01 mm day$^{-1}$ (with respect to the baseline values given by [98]) and the new daily $ET_{0,j}$ is determined. For the sake of simplicity, we discuss here only the changes in virtual water content associated to positive variations of $ET_{0,m}$.
Depending on the soil water content and irrigation conditions, the new $ET$ demand can be totally or partially satisfied. In the irrigated production, the new water requirement can be partly met by irrigation, with larger evapotranspirations of blue water. In the rainfed production, the ability of the new evapotranspiration demand to be met depends on the water available from precipitation. Higher evapotranspiration demand can take better advantage of precipitation (i.e., higher evapotranspiration of green water), if available, thereby limiting runoff and water losses. The $CWF$ variations have opposite directions for different crops as shown by the sensitivity indexes in Figure 2.9. Soybean exhibits a positive sensitivity index around 0.08, indicating that a positive variation of $ET_{0,m}$ increases the crop water footprint. A possible reason lies in the yield response factor, $K_y$ – that relates $Y$ reductions to $ET$ reductions in equation (3.5) –, which is equal to 0.85. According to [38], $K_y < 1$ implies that the crop exhibits a less-than-proportional increase in the yield with increased actual evapotranspiration. Figure2.10 shows the $SI_{ET_{0,m}}$ spatial variability with a high resolution map. Considering the biggest producers, the United States (located in the temperate belt) exhibits lower sensitivities to $ET_{0,m}$ variations than Brazil and India (located in the tropical zone), where the $SI_{ET_{0,m}}$ reaches values around 0.15. Wheat, rice, and maize show negative sensitivity indexes; for these crops, an increased evapotranspiration reduces the virtual water content due to increased yields. These crops, in fact, are very

sensitive to water surplus, as shown by their yield response factor which is equal or higher than 1, indicating that the yield increases more-than-proportionally when $ET$ increases. The $CWF$ of rice is the most sensitive to evapotranspiration variations, as shown by a $SI_{ET_{0,m}}$ value of -0.4 on average. The sensitivity index of rice is quite heterogeneous at sub-national scale (Figure 2.11), with values between -0.10 and-0.60 (e.g., South Sudan, Ethiopia, Tanzania).

### 2.4.3 Length of the growing season

The length of the growing period, $LGP$, defined by the planting and the harvesting date, is used to calculate the length of the 4 growth stages ($l_{st}$) defining the shape of the piecewise crop coefficient curve.

The sensitivity of $CWF$ to $LGP$ is evaluated by varying $LGP$ of $\pm 1$ day. The variation makes a stage of the growing season 1 day longer or shorter than the nominal value, while the other 3 stages are shifted of one day (i.e., the crop is harvested later or earlier), maintaining their initial length. Such translation changes the daily crop water requirement because $k_c$ is differently associated with daily $ET_0$ values, impacting the virtual water content.

Wheat exhibits a sensitivity index of -0.05 on average (see Figure 2.9). The negative values of $SI_{LGP}$ indicate that the virtual water content has decreased due to the increased yields. In fact, a one day longer growing season implies a higher $ET$ demand, which can be partially or totally met by precipitation or irrigation, depending on the cultivation conditions. The spatial variability of the wheat $SI_{LGP}$ (Figure 2.12) ranges between the tropical zone where the sensitivity indexes are around -0.05 and the temperate belt and the sub-tropical (summer rainfall) zones where these indexes reach values around -0.25. In these areas, in fact, the yield increases more than elsewhere (about 1.5-2.5% with respect to the baseline value) because the increased $ET$ demand is totally met by irrigation in the Nile Basin, in the North of India, and in the North-East of China, and by precipitation in Belgium, Netherlands, Northern Italy, and Mongolia. Rice, maize, and soybean show negative sensitivity indexes of -0.18, -0.09, -0.03, respectively.

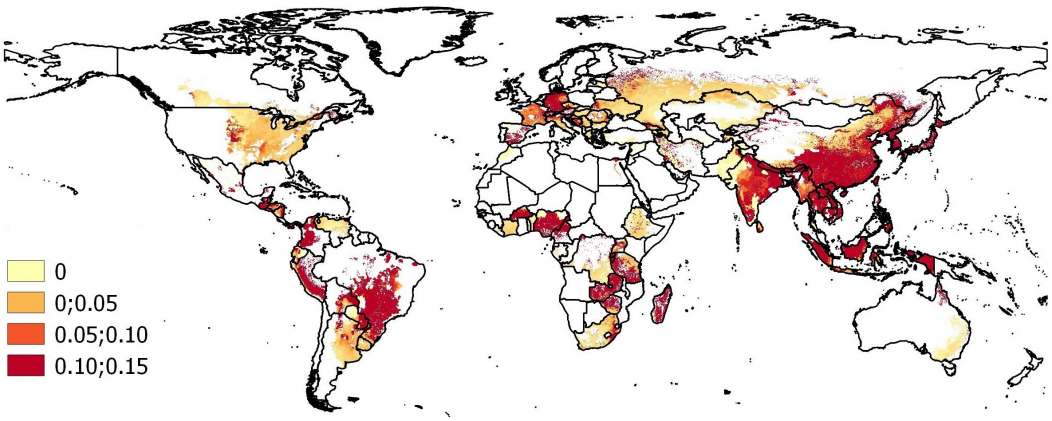


Fig. 2.10 The sensitivity index of the $CWF$ of soybean to the reference evapotranspiration ($ET_0$). $ET_0$ is varied of 0.01 mm day$^{-1}$.

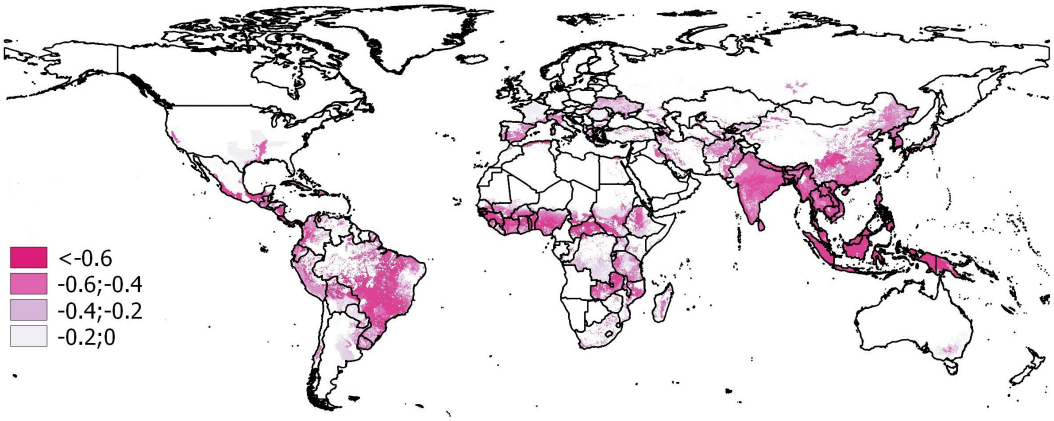


Fig. 2.11 The sensitivity index of the $CWF$ of rice to the reference evapotranspiration ($ET_0$). $ET_0$ is varied of 0.01 mm day$^{-1}$.

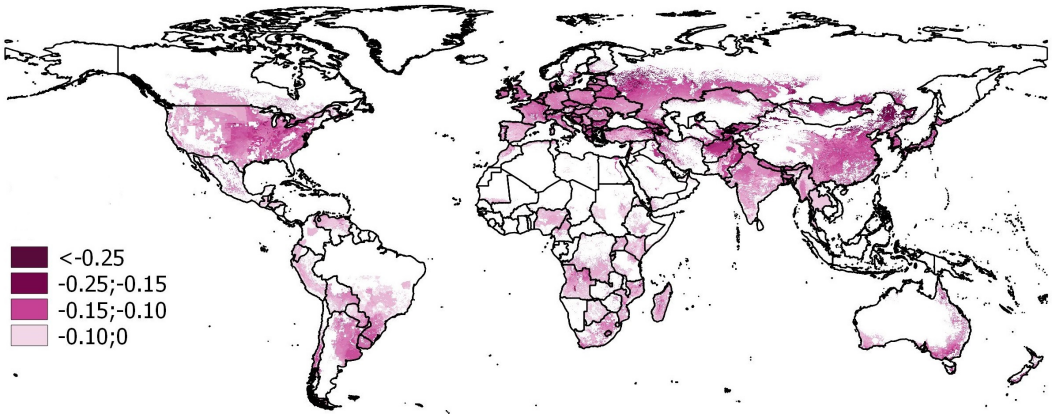


Fig. 2.12 The sensitivity index of the $CWF$ of wheat to the length of the growing period ($LGP$). The length of the growing period is varied of 1 day.

### 2.4.4 Crop planting date

Crop planting dates were taken from [108]. This database provides the months when the growing season starts and ends, making a distinction between rainfed and irrigated production. The intermediate day of the month is taken as the planting date. Varying the planting date (±1 day), with constant length of the growing period, implies a rigid translation of the growing season to higher or lower daily reference evapotranspirations. Therefore, the $SI_{PD}$ depends on the month of the year in which the crop is planted and on the temporal evolution of $ET_0$ during the growing season, and it may also include negative values. On global average, rice, maize, and soybean exhibit positive sensitivity indexes: 0.08, 0.18, 0.1, respectively, with cells showing strong spatial heterogeneity, especially for rice harvested areas where some cell show negative sensitivity indexes (see Figure 2.9). For these crops, a 1 day shift in the growing period increases the virtual water content. Rice, maize, and soybean are, in fact, spring or summer crops, thus a positive variation of the planting date implies a shift of the growing period to lower reference evapotranspiration periods, and thus lower crop water requirement and virtual water content. The map in Figure 2.14 better specifies the spatial variability shown by the boxplots; for example, focusing on the main rice producers (e.g., China, India, Vietnam), the $SI_{PD}$ values vary from -1 to 1 depending on the water conditions of each harvested area. Southern India shows a $SI_{PD}$ value of -1 indicating an attenuation of the crop water footprint due to the increased yield of 0.5%; Northern India exhibits a positive sensitivity index around 0.8 due to a decreased yield, which reaches -2% in those cells under rainfed conditions. The map in Figure 2.15 shows the $SI_{PD}$ values of maize and help to localize the positive $SI_{PD}$ values; the temperate zone is positively sensitive to planting date changes due to yield reductions caused by $ET$ attenuations. The $SI_{PD}$ value presents a significant within-countries spatial variability, which is particularly evident in Brazil, India, and China. The sensitivity index of wheat $CWF$ goes in the opposite direction with an average value of -0.05. This means, that delaying the planting date of wheat (which is mostly a winter crop) helps to reduce its water footprint. However, in some wheat producing areas (located in the tropical belt for example) the behaviour is the opposite one, with positive $SI_{PD}$ values.

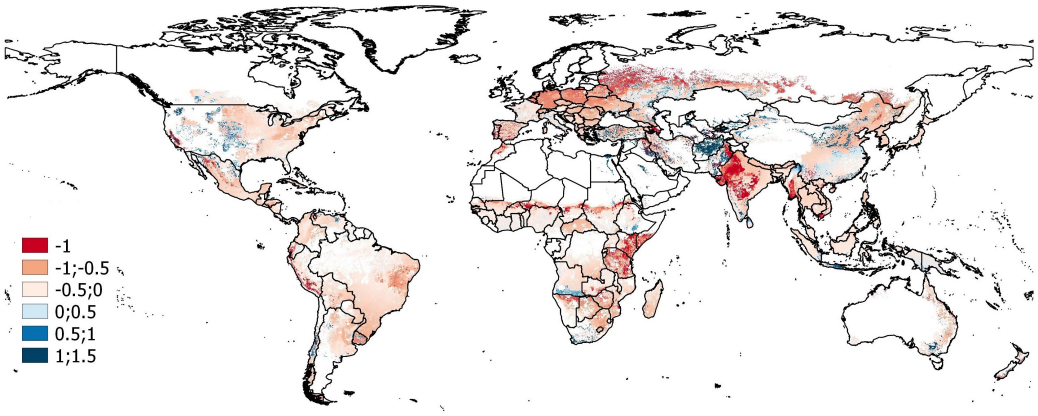


Fig. 2.13 The sensitivity index of the $CWF$ of maize to the length of the growing period ($LGP$). The length of the growing period is varied of 1 day.

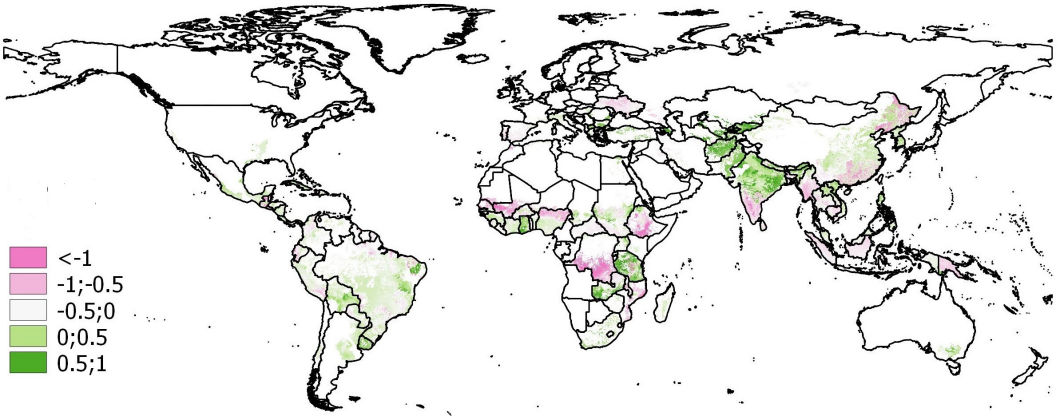


Fig. 2.14 The sensitivity index of the $CWF$ of rice to the planting date ($PD$). The planting date is varied of 1 day.

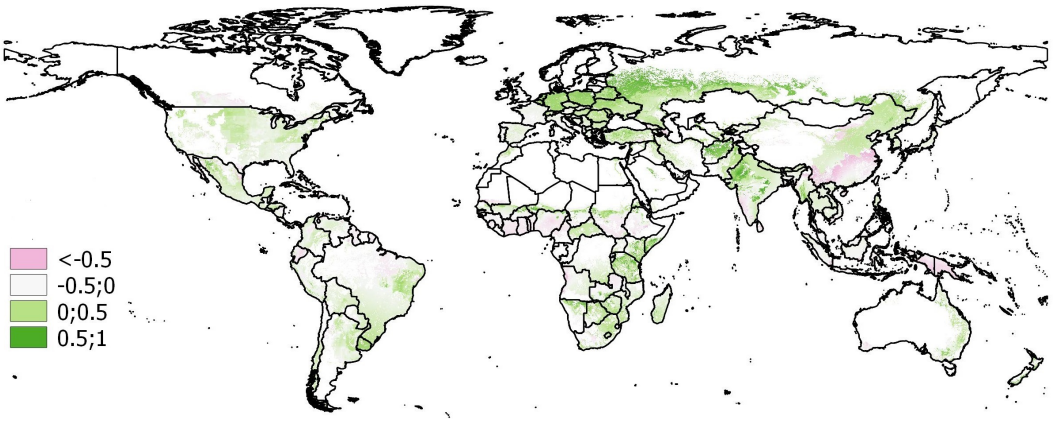


Fig. 2.15 The sensitivity index of the $CWF$ of maize to the planting date ($PD$). The planting date is varied of 1 day.

## 2.5 Concluding remarks

The high resolution maps of the crop water footprint $CWF$ of the main cultivated crops have been obtained using recently high-resolution data and accounting for multi-cropping practices, thus taking into account precipitation and temperature variabilities along the year, and also accounting for different irrigation requirements. The $CWF$ values differ substantially among crops and across production regions, exhibiting strong spatial heterogeneity even at the sub-national scale. The spatial heterogeneity in the $VWC$ is mainly driven by the yield patterns with a correlation coefficient higher than 0.7 for all the crops. This suggests that the crop $VWC$ is influenced by agricultural practices more than climatic conditions. This is an important result, especially in terms of indicating a strategy toward virtual water content reduction; [113], for example, have shown that there is a great opportunity to improve water productivity (and thus reduce $CWF$) through the improvement of yield levels within the available water balance in rainfed agriculture, without requiring additional blue water resources. Furthermore, considering the logarithmic relationship existing between $CWF$ and yield, the largest water productivity gains can be achieved in the two main hot-spot regions of the world in terms of poverty and water scarcity, i.e. sub-Saharan Africa and South Asia, where the $CWF$ estimates are generally above the global average due to the low crop yields.

Aggregate analyses at the continental scale provide a global view of the $CWF$ value in relation to crop production. From the histograms in Figure 2.7 and from the boxplots in Figure 2.8, it is clear that wheat, rice, and maize are characterized by a higher water productivity (and thus lower virtual water content) than soybean, due to their higher yields. However, soybean yield, as well as the production area, is expected to increase with the help of genetic resources which may provide the solution needed to overcome abiotic and biotic constraints [65]. The results of the aggregate analysis also show the global consumptive water use of the four grains, which is about 3300 $km^3$ $yr^{-1}$ in the period from 1996 to 2005. Green water contributed to 90% of the global consumptive water use in the crop growing periods; this high proportion of green water is partly due to the dominance of rainfed agriculture. In addition, in irrigated lands, green water contributed to 25%-80% of the total consumptive water use as also shown, at the grid cell scale, by the maps of blue $CWF$. In

fact, only in some regions and countries (e.g., Egypt, Pakistan, Saudi Arabia) crop production depends primarily on blue water. The important role of green water in crop production highlights the need for a better management of this water resource.

Most notably, to our knowledge this is the first study assessing, at the global scale, the sensitivity of the $CWF$ estimates to the model-inputs. The results of the sensitivity analysis show that wheat is the most sensitive crop to the length of the growing period, rice to the reference evapotranspiration, maize and soybean to the crop planting date. Virtual water content shows different sensitivity to input parameters not only among crops, but also across the harvested areas of the world, even at the sub-national scale. These results may inform future efforts aiming at the refinement of data used in the assessment of agricultural water requirements and lend themselves to the identification of the parameters that farmers and land managers can modify to effectively reduce the water cost for crop production. Crop water footprint estimates and sensitivity studies will need to be extended towards other crops and other water using processes, at different spatio-temporal scales, to have a complete picture of this effective tool to tackle water and food security.

# Chapter 3

# A Fast-Track approach to deal with temporal dimension of CWF

*The work described in this chapter has been partially derived from paper [138].* Population growth, socio-economic development and climate changes are placing increasing pressure on water resources. Crop water footprint is a key indicator in the quantification of such pressure. It is determined by crop evapotranspiration and yield value, which can be highly variable in space, as shown in Chapter 2, and time providing the significant trends that have been shown for yields [111, 110]. While a great deal of attention has been devoted to the $CWF$ variability in space, less attention has been paid to its variability in time. To date, only local studies have evaluated a time-varying crop water footprint [128, 149, 109], with particular regard to the Chinese case [129, 155, 154]. Recently, a number of studies have adopted a simple approach that ascribes the time variability of crop water footprint only to yield trends, leaving out the effects of evapotranspiration variations [81, 80, 31, 32]. However, the feasibility of this approach, we call "Fast-Track approach", has yet to be proved. This approach has been adopted bothe for local and global studies, but its feasibility is yet to be provided. Can this approach capture the main $CWF$ temporal variability? How big is the error arising with the assumption of constant evapotranspiration? In the first part of the Chapter, the Fast-Track method is described and validated. In the second part of the Chapter, the uncertainty of

the method is assessed and an example of application is provided, pertaining the virtual water trade of wheat, maize, rice, and soybean. Results show and confirm the hypothesis that water footprint changes are mainly driven by yield trends. The error associated with the Fast-Track method due to considering constant evapotranspiration is three-times smaller than the uncertainty of the model used to compute the crop water footprint.

## 3.1 The Fast-Track approach: assumption and validation

Recent literature on virtual water testifies a growing application of a Fast-Track (FT) approach for introducing the time dependency in crop water footprint assessment, with the main objective of calculating the volumes of virtual water embedded in internationally-traded agricultural goods.
According to the FT approach, the crop water footprint of country $c$ in year $t$, $CWF_{c,t}(Y)$, is only driven by crop yield variations, $Y_{c,t}$ [ton·ha$^{-1}$], while evapotranspiration depth, $\overline{ET}_{c,T}$ [mm], is kept constant to an average value typical of a reference year or period ($T$), namely

$$CWF_{c,t}(Y) = \frac{10 \cdot \overline{ET}_{c,T}}{Y_{c,t}} \qquad \left[\frac{m^3}{ton}\right], \tag{3.1}$$

where, 10 is a corrective factor to convert the evapotranspiration depth from mm to m$^3$·ha$^{-1}$. With this formulation of time-varying $CWF$, it is implicitly assumed that the variations of crop evapotranspiration have negligible effects on the crop water footprint when compared to the effects of yield variations and thus the $\overline{ET}_{c,T}$ value can be fixed for any year $t$. The advantage behind equation (3.1) is that yield time-series data are easily available at the country scale (e.g., FAOSTAT database), and thus the $CWF$ variability can be obtained without the adoption of computational-demanding models that are generally used to estimate evapotranspiration. Equation (3.1) has been adopted in previous studies to include time variations in the analyses of virtual water trade [81, 80, 31, 32], but without testing the suitability nor the uncertainty of the

adopted methodology. Validation of the FT approach is one of the purpose of this Chapter.

The FT approach allows one to exploit average crop water footprint estimates previously determined over a period $T$, $\overline{CWF}_{c,T}$. Literature accounts a number of $CWF$ estimates at different spatial scale and averaged over different time-intervals [126, 89, 136]. These time-averaged crop water footprints can be scaled through yield trends following

$$CWF_{c,t}(Y) = \frac{\overline{CWF}_{c,T} \cdot \overline{Y}_{c,T}}{Y_{c,t}} \qquad \left[\frac{m^3}{ton}\right], \tag{3.2}$$

in order to make them time-dependent. $\overline{Y}_{c,T}$ is the average crop yield over $T$ while $Y_{c,t}$ is the country-yield of year $t$. Equation (3.2) has been recently applied by *Duarte et al.* [39] to compute annual virtual water flows from 1965 to 2010 for 133 products.
To date, equations (3.1,3.2) have been applied only at the country scale. However, they can be applied at any spatial resolution, depending on the goals and data availability. Thus, symbol $c$ can refer also to a region, a province or a cell and the time-interval $T$ can indicate both a single year or a temporal window of two or more years length.

### 3.1.1 Validation of the Fast-Track approach

Here we test and validate the assumption of constant evapotranspiration that grounds the FT approach. The aim of validation is twofold: (i) to support previous studies that have applied the method without examining in depth its feasibility and (ii) to foster its adoption to deal with temporal variability in future water footprint assessment. In order to test the method, we compare the $CWF$ estimates obtained with the FT approach with the estimates accomplished through a more refined model accounting for both the inter-annual yield values and the evapotranspiration changes. The two different estimates are obtained as detailed in the following for wheat, rice, maize, and soybean. These crops provide more than 50% of the global caloric content of human diet [34], they contribute for more than 50% to the global water footprint [89] and they account for over 30% of the global virtual water trade of agricultural goods

[132]. The validation could be accomplished for any other product, provided that data are available (see below).

### 3.1.2 Evaluation of the crop water footprint through the FT approach

Annual $CWF$ estimates are carried out according to equation (3.2) applied at the country scale for the period 1961-2013. Equation (3.2) requires as input the average crop water footprint over a reference period $T$ ($\overline{CWF}_{c,T}$), the average yield values over $T$ ($\overline{Y}_{c,T}$), and the annual yield data from 1961 to 2013 ($Y_{c,t}$). The average crop water footprint values are provided by *Tuninetti et al.* [136] for the period $T$=1996-2005 at 5x5 arc minute resolution. To obtain country-averages, these gridded estimates are aggregated through a production-weighted mean (see *Tuninetti et al.* for further details). The country-yield averages $\overline{Y}_{c,T}$ are obtained by averaging the annual FAOSTAT data available for each producing-country from 1996 to 2005; finally, the annual country-yield values $Y_{c,t}$ are derived from the same database with $t$ running from 1961 to 2013. We remark that the annual $CWF_{c,t}(Y)$ estimates obtained with the FT approach are easy and fast to be computed once the $\overline{CWF}_{c,T}$ values are known.

### 3.1.3 Evaluation of the crop water footprint with the detailed method

The $CWF_{c,t}(Y)$ estimates obtained with the FT approach are compared with the annual water footprint estimates achieved when both the yield and the evapotranspiration changes are taken into account. To this purpose, we adapted equation (2.1) used for time-fixed assessments, by introducing the time variability of both yield and evapotranspiration.
The yield-and evapotranspiration-dependent annual crop water footprint in cell $i$ of year $t$ belonging to the range 1961:2013, $CWF_{i,t}(Y, ET)$, reads

$$CWF_{i,t}(Y, ET) = \frac{10 \cdot ET_{i,t}}{Y_{i,t}} \qquad \left[\frac{m^3}{ton}\right]. \tag{3.3}$$

In this case, both the crop evapotranspiration, $ET_{i,t}$, and the crop yield, $Y_{i,t}$, are time-dependent, differently from equations (3.1,3.2) where $ET$ values are assumed constant and averaged over $T$.
The annual $ET_{i,t}$ value is the water depth actually evapotranspired by the crop during the growing season of year $t$. It is determined [8] as the product between the potential evapotranspiration ($ET_0$), a crop coefficient (which is characteristic of the crop height, canopy resistance, and soil evaporation rate), and a water stress coefficient obtained through a daily water balance (see Section 1.1.2). We assume that crop properties (e.g., planting date, length of the growing period) and soil characteristics (e.g., available soil water content) remain constant along the study period due to lack of more detailed data. Differently, we account for inter-annual fluctuations of potential evapotranspiration and precipitation integrating the annual climatic data provided by the CRU database [140] and the GAEZ database [98]. The CRU database covers the period between 1961 and 2013 providing for each year gridded potential evapotranspiration and precipitation at 30x30 arc minute resolution on monthly basis. The values given by the GAEZ database cover the period between 1961 and 2000 with yearly temporal resolution on a 5'x5' grid. The combination of the two databases allows one to achieve the best spatio-temporal resolution in the estimation of the $ET_0$ values.
For the crop yield, time series of gridded yield data are not available at the spatial resolution required by equation (3.3) for the period 1961-2013. Therefore, in order to obtain time-variable gridded data, we adjust the values provided by *Monfreda et al.* [94] at 5'x5' resolution for year $t$=2000, i.e., $Y^{Mo}_{i,t=2000}$ with two factors, namely

$$Y_{i,t} = \alpha^{cl}_{i,t} \cdot \alpha^{man}_{c,t} \cdot Y^{Mo}_{i,t=2000} \qquad \left[\frac{ton}{ha}\right]. \tag{3.4}$$

The factor $\alpha^{cl}_{i,t}$ accounts for climate-driven yield changes while the factor $\alpha^{man}_{c,t}$ accounts for the yield changes induced by technological advances and agricultural improvements, ascribable to the anthropic (man) role in agriculture. Depending on data availability, $\alpha^{cl}_{i,t}$ can be defined at the cell level while $\alpha^{man}_{c,t}$ can only be defined at the country scale.
The factor $\alpha^{cl}_{i,t}$ accounts for yearly fluctuations of crop yield at the cell level, due

to year-to-year changes in crop evapotranspiration. Such changes are assumed to impact the yield according to the relation proposed by *Doorenbos et al.* [38],

$$1 - \frac{Y^{cl}_{i,t}}{Y_{i,t=2000}} = k_y \cdot \left(1 - \frac{ET_{i,t}}{ET_{i,t=2000}}\right), \tag{3.5}$$

where, $Y^{cl}_{i,t}$ is the yield in year $t$ when only variations in crop evapotranspiration are considered. Thus, *cl* marks the new yield determined by climatic changes only. Equation (3.5) relates the relative change in evapotranspiration to the relative change in crop yield through the yield response factor, $k_y$ [38]. We refer the changes to year $t$=2000 because the yield dataset by *Monfreda et al.*, $Y^{Mo}_{i,t=2000}$, is representative for that year. The value of $\alpha^{cl}_{i,t}$ is determined by equations (3.4) and (3.5) assuming $\alpha^{man}_{c,t} = 1$ and thus $Y_{i,t} = Y^{cl}_{i,t}$, namely

$$\alpha^{cl}_{i,t} = 1 - k_y \cdot \left(1 - \frac{ET_{i,t}}{ET_{i,t=2000}}\right). \tag{3.6}$$

When only climatic variations are taken into account, the yield value reads

$$Y^{cl}_{i,t} = \alpha^{cl}_{i,t} \cdot Y^{Mo}_{i,t=2000}, \tag{3.7}$$

Gridded yield values obtained with equation (3.7) are then aggregated at the country scale through a weighted mean, i.e.,

$$Y^{cl}_{c,t} = \frac{\sum_{i \in c} Y^{cl}_{i,t} \cdot A_{i,t=2000}}{\sum_{i \in c} A_{i,t=2000}}, \tag{3.8}$$

using the gridded harvested area of year 2000, $A_{i,t=2000}$, provided by *Portmann et al.* [108] as the weights. These country-values are used in the following to determine the $\alpha^{man}_{c,t}$ factor.

The $\alpha^{man}_{c,t}$ factor expresses the yield variability due technological and mechanical advances in the agricultural management (e.g., use of pesticides, application of fertilizers, extensive irrigation). Essentially, $\alpha^{man}_{c,t}$ is thought as a correction factor to the $Y^{cl}_{c,t}$ values in order to account for all other aspects beyond climate. It is defined as the ratio between the FAO country-scale yield, $Y^{FAO}_{c,t}$, and the national $Y^{cl}_{c,t}$ values calculated with equation (3.8)

$$\alpha_{c,t}^{man} = \frac{Y_{c,t}^{FAO}}{Y_{c,t}^{cl}}. \tag{3.9}$$

With the adoption of equations (3.3,3.4) it is now possible to determine the annual crop water footprint in each cell. Country-estimates, $CWF_{c,t}(Y, ET)$, are then obtained through a production-weighted mean of the gridded values, where cell production is given by the product between the $Y_{i,t}$ values (expressed in ton·ha$^{-1}$) and the harvested area $A_{i,t=2000}$ (in ha) provided by *Portmann et al.* [108].

### 3.1.4 Comparison between the two methodologies

In Figure 3.1 we compare annual $CWF$ estimates achieved through the Fast-Track approach ($CWF_{c,t}(Y)$), only accounting for the yield variability, with those accomplished by the detailed method ($CWF_{c,t}(Y, ET)$) accounting not only for yield but also for the evapotranspiration variability. Each point in the scatter represents the national crop water footprint in year $t$ within the period 1961-2013. The estimates obtained with the two approaches compare well for all crops: in fact all points are mostly aligned along the 1:1 line with limited scatter, as confirmed by the values of the coefficient of determination, $R^2$ (that read 0.977 for wheat, 0.965 for rice, 0.973 for maize, and 0.914 for soybean). The overall agreement between the two methods confirms that the temporal variability of the crop water footprint is mainly driven by yield variations, while the variability of crop evapotranspiration, that is kept constant over time in the FT method (and not in the refined method), appears to play a negligible role. We remark that this does not correspond to neglect the relevance of the climatic variations on the $CWF$: in fact, the climatic signature remains in the yield time series. For example, *Ray et al.* ([110] have found that around 30% of the wheat, rice, maize and soybean yield variability is explained by climate variability through the inter-annual fluctuations of precipitation and temperature values.

Moreover, the FT method performs well independently of the presence of yield trends. In fact, there are countries in the database where yield has improved over time inducing the decrease in $CWF$; whereas, in other countries, yield has stagnated or decreased, making the $CWF$ values remain constant or increase.

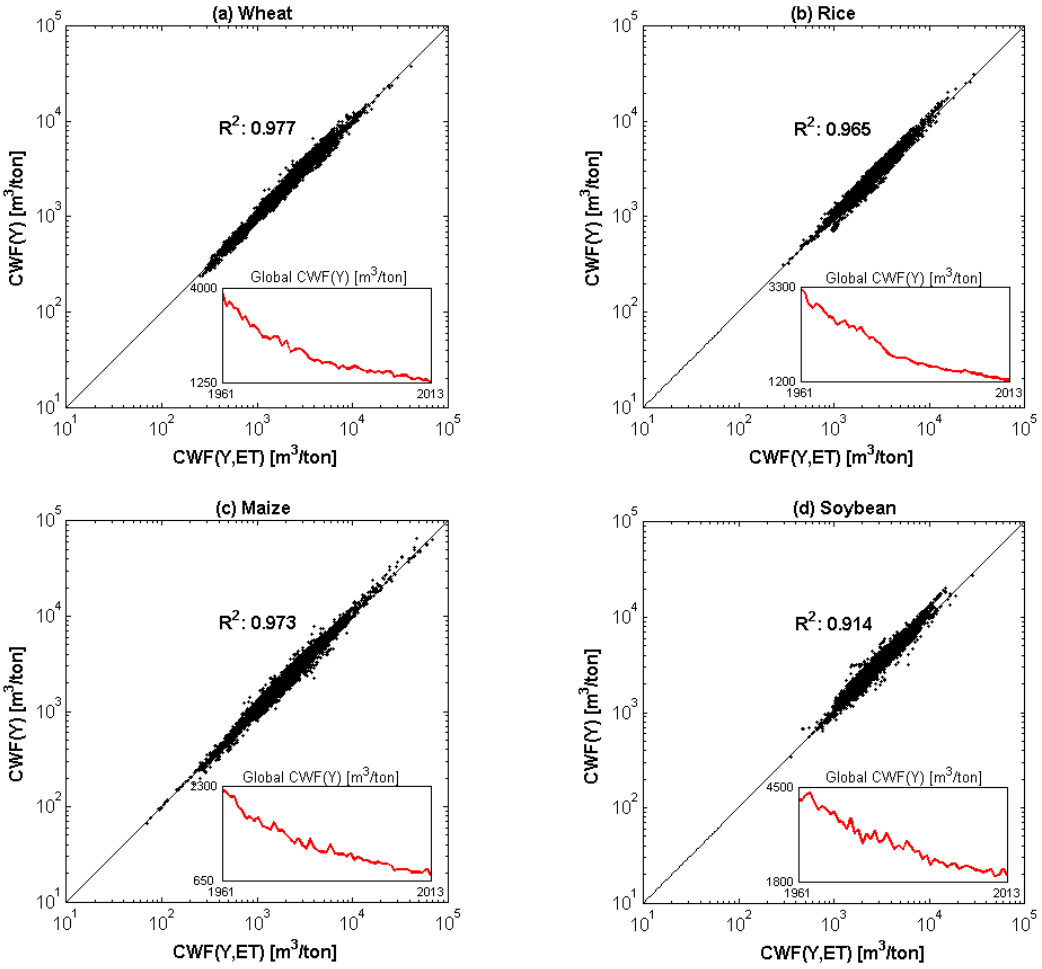


Fig. 3.1 Comparison of the annual crop water footprint ($CWF$) estimates obtained by the Fast-Track approach, $CWF(Y)$, with the values obtained with the detailed methodology accounting for both yield and evapotranspiration variations, $CWF(Y, ET)$. The comparison is made at the country scale across the period between 1961 and 2013 for wheat (a), rice (b), maize (c), and soybean (d). The $R^2$ value indicates the coefficient of determination between the two estimates. The inset of each panel reports the global trend of the average crop water footprint evaluated with the FT method. Such annual global value is obtained through a weighted mean of the country-estimates, using the annual country production as the weight.

According to *Ray et al.* [111], wheat, rice, maize, and soybean are experiencing yield increases in around 70% of their harvested areas, resulting in a decrease of crop water footprints, stagnation in over 20% of the areas and collapse in the remaining areas. Despite the strong spatial heterogeneity of $CWF$ trends worldwide, the global average water footprint of each crop has sharply decreased from 1961 to 2013, as shown by the red lines in the insets of Figure 3.1.

## 3.2 Uncertainty of the FT approach

The uncertainty of the FT approach is now assessed and decomposed in its main components. Denoting the real (unknown) crop water footprint of country $c$ in year $t$ as $CWF^r_{c,t}$, the error structure is here assumed to be multiplicative to account for the fact that crop water footprint is positive-valued, namely

$$CWF_{c,t}(Y) = CWF^r_{c,t} \cdot \epsilon_{c,T} \cdot \epsilon'_{c,t}. \tag{3.10}$$

The $\epsilon_{c,T}$ error is due to the type of model adopted to calculate the crop water footprint; it impacts the $\overline{ET}$ value in equation (3.1) and the $\overline{CWF}$ value in equation (3.2). The $\epsilon'_{c,t}$ error arises from the assumption of constant evapotranspiration in the FT approach.
The $\epsilon_{c,T}$ error can arise for different reasons, depending on the model and data used to estimate the crop evapotranspiration (e.g., the data regarding cultivated and irrigated areas, growing periods, crop parameters, soil, climate), and the yield data. In order to quantify such error, we compare the average $CWF$ estimates obtained in Chapter 1 and already used in equation (3.2) at the country scale, with the country-estimates given by *Mekonnen et al.* [89] which constitutes the overriding reference study in the literature of water footprint assessment. Both estimations are referred to the period $T$=1996-2005; we denote as $\overline{CWF}^{Me}_{c,T}$ the estimates by *Mekonnen et al.* and $\overline{CWF}^{Tu}_{c,T}$ the values derived from Chapter 1.
We calculate, for each country and for each crop, the corresponding $\epsilon_{c,T}$ error, as

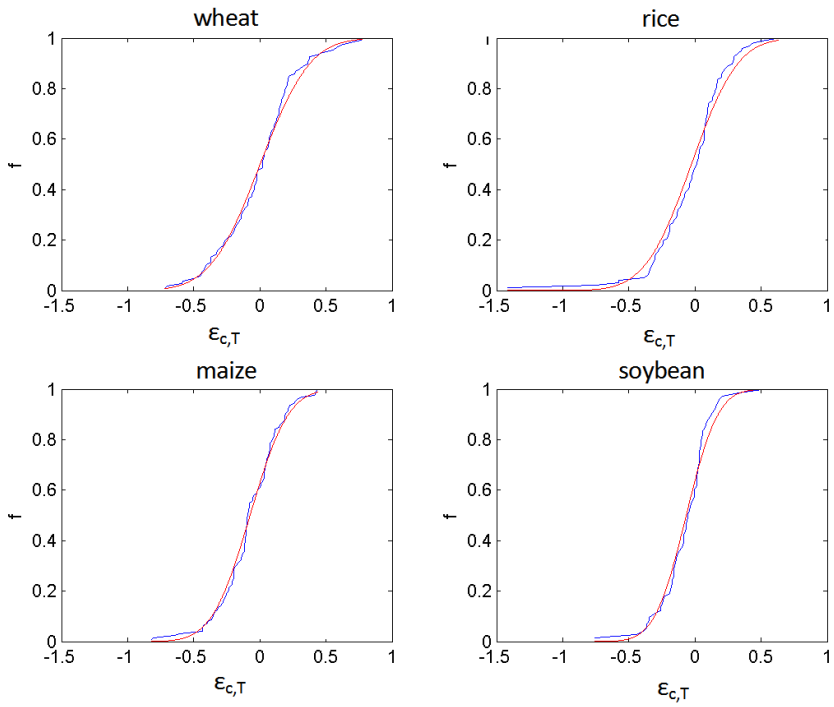


Fig. 3.2 Frequency distribution of the empirical error associated with the model adopted to compute the crop water footprint, i.e., $\epsilon$, (blue line) and fitted log-normal distribution (red curve).

$$\epsilon_{c,T} = \frac{\overline{CWF}^{Tu}_{c,T}}{\overline{CWF}^{Me}_{c,T}}, \tag{3.11}$$

We thus obtain four samples of $\epsilon_{c,T}$ values, one for each crop (the length of each sample is reported in Table 3.1). We find that each sample is fitted by a two-parameter log-normal distribution (see Figure 3.2), with parameters $\mu$ and $\sigma$ representing the average and standard deviation of the log-transformed data, given in Table 3.1.

Overall, $\mu$ is around 0 for all crops while $\sigma$ is between 0.25 and 0.30. These relatively large $\sigma$ values imply a high sensitivity of the crop water footprint to the model parameters and input data used, as previously shown in other studies [89, 153].

In Figure 3.3 we compare the $\overline{CWF}^{Tu}_{c,T}$ and $\overline{CWF}^{Me}_{c,T}$ estimates; each circle represents a producing-country and the size of the circle indicates the share of the country in the global production. The largest producer of each crop

Table 3.1 Statistics of the error, $\epsilon$, associated to the methodology described by *Tuninetti et al.* [136] and statistics of the error, $\epsilon'$, associated to the FT method assumption of invariable evapotranspiration. The $l(\epsilon)$ and $l(\epsilon')$ values indicate the length of the error samples available for each crop.

| | $\log(\epsilon)$ | | | $\log(\epsilon')$ | | |
|---|---|---|---|---|---|---|
| | $l(\epsilon)$ | $\mu_\epsilon$ | $\sigma_\epsilon$ | $l(\epsilon')$ | $\mu_{\epsilon'}$ | $\sigma_{\epsilon'}$ |
| wheat | 5689 | -0.001 | 0.296 | 107 | -0.022 | 0.093 |
| rice | 5405 | 0.012 | 0.286 | 97 | -0.016 | 0.099 |
| maize | 6958 | 0.036 | 0.266 | 126 | -0.066 | 0.104 |
| soybean | 3680 | 0.041 | 0.254 | 73 | -0.086 | 0.135 |

is highlighted by a red circle. Generally, the estimates compare well for all crops with average coefficients of determination, $R^2$, always higher than 0.7. However, when weighted by country production the $R^2_w$ values suggest better or worse agreement between the estimates provided by *Tuninetti et al.* and *Mekonnen et al.* depending on the crop. For rice and maize (panels (b,c)), the agreement between the two studies is particularly high, with $R^2_w$ equal to 0.89 and 0.83, respectively. Conversely, for wheat and soybean the $R^2_w$ values are lower, particularly for soybean ($R^2_w$=0.41).

The $\epsilon'_{c,t}$ error is determined as the ratio between the $CWF_{c,t}(Y)$ values, estimated with the Fast-Track approach according to equation (3.2), and the $CWF_{c,t}(Y, ET)$ values achieved with the refined method, i.e.,

$$\epsilon'_{c,t} = \frac{CWF_{c,t}(Y)}{CWF_{c,t}(Y, ET)}. \tag{3.12}$$

As for the $\epsilon_{c,T}$ errors, we find that the $\epsilon'_{c,t}$ values follow a log-normal distribution (see Figure 3.4); the $\mu$ and $\sigma$ values are shown in Table 3.1 together with the length of the $\epsilon'_{c,t}$ samples. For all crops, the precision of the estimates is high, with a standard deviation of the error around 0.1, confirming the good agreement between the two estimators previously shown in Figure 3.1.
The uncertainty in the annual $CWF$ estimates ascribable to the assumption of constant evapotranspiration (in the FT approach) results three-times lower than the model uncertainty, evaluated as a comparison between the outcomes provided by *Mekonnen et al.* and those derived from *Tuninetti et al.* . Therefore, the FT approach is appropriate to deal with the time variability of crop water footprint.

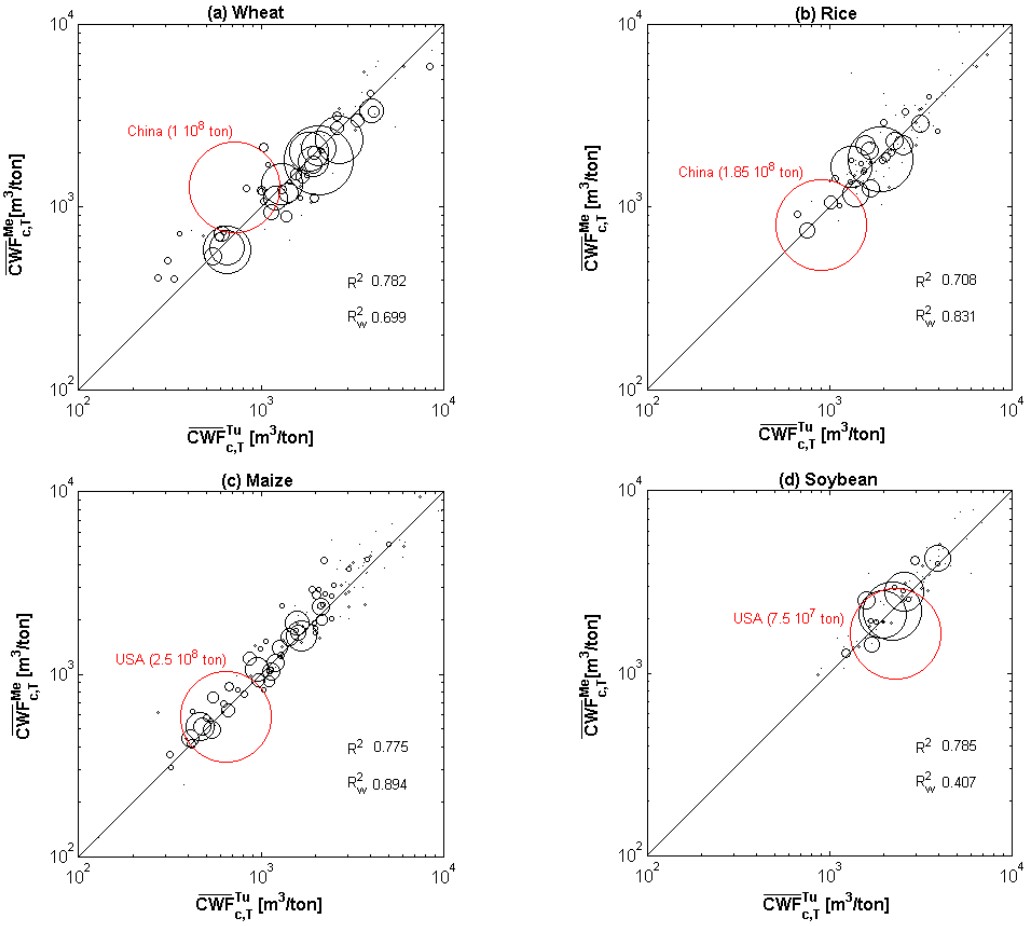


Fig. 3.3 Comparison of the average crop water footprint as estimated with the methodology provided by *Tuninetti et al.* [136], $\overline{CWF}^{Tu}_{c,T}$, with the values taken from *Mekonnen et al.* [89], $\overline{CWF}^{Me}_{c,T}$, for wheat (a), rice (b), maize (c), and soybean (d). The area of each circle in the graph is proportional to the share of the country to the global annual production, while the red circles highlight the greatest producing countries. $R^2$ indicates the overall coefficient of determination and $R^2_w$ stands for the coefficient of determination weighted by countries annual production along the time-window of interest.

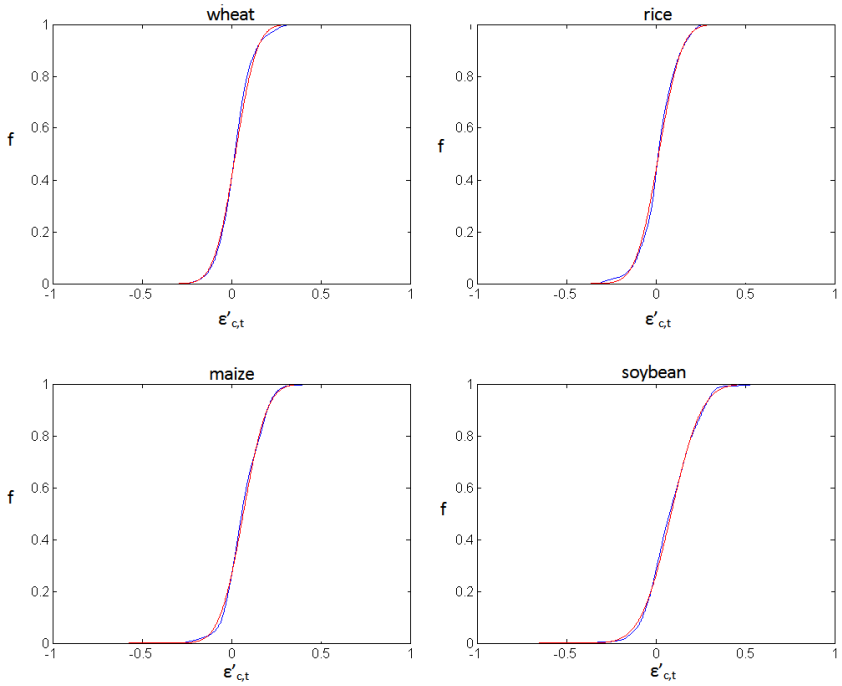


Fig. 3.4 Frequency distribution of the empirical error associated with the Fast Track approach, i.e., $\epsilon'$, (blue line) and fitted log-normal distribution (red curve).

## 3.3 Example of application: the case of virtual water trade

The time-dependent $CWF_{c,t}(Y)$ estimates, obtained for wheat, rice, maize, and soybean with the FT approach, are now used to assess the temporal variations of the virtual water volumes embedded in the international trade. To this aim, we calculate the annual virtual water embedded in each crop exported by country $c$ in year $t$, $VW_{c,t}$, as the product between the weight of crop, $W_{c,t}$, (in tonnes) exported by country $c$ and the annual water footprint of the crop, $CWF_{c,t}(Y)$, for the period between 1986 and 2011. The $W_{c,t}$ values are available from the FAOSTAT database, whereas the crop water footprint values have been estimated by equation (3.2). The total virtual water trade, $VWT_t$, is then built by summing up the $VW_{c,t}$ of all crops and countries, and shown by the solid line in Figure 3.5. During the period 1986-2011 countries have been moving growing volumes of virtual water, embedded in the four study

crops, worldwide: from 300 km$^3$ in 1986 to 540 km$^3$ in 2011 (refer to the solid line, Figure 3.5).

In order to provide evidence of the importance of using time-dependent $CWF$ values, we report in the same graph the annual virtual water trade data obtained using the annual trade from the FAOSTAT database and the average $\overline{CWF}_{c,T}^{Tu}$ values over the period 1996-2005 taken from *Tuninetti et al.* [136]. In this case, the virtual water content of each crop is kept constant over time and the $VWT$ trend is only driven by the amount of products that are internationally-exchanged over time, i.e., $W_{c,t}$. We observe significant differences between the trend obtained with time-variable $CWF_{c,t}(Y)$ and the time-averaged $\overline{CWF}_{c,T}^{Tu}$ virtual water content: e.g., in year 2011 the difference is around 100 km$^3$. Such comparison exemplifies for the four study crops the gap existing in the VW trade estimations between the two approaches.

Finally, the green area in Figure 3.5 depicts the 90% confidence interval of the $VWT$ estimation. The confidence interval is determined as $VWT_t \pm z^* \cdot \sigma_{VWT}$, where $z^*$ is the $95th$ percentile of a standard normal variate and $\sigma_{VWT}$ is the standard deviations of the total virtual water flow. The square $\sigma^2_{VWT}$ value is equal to the sum of the variance associated to the virtual water trade of each crop $cr$ (assuming independence of the four virtual water flows). Such variance is calculated as the product among (i) the variance of the $\epsilon'$ errors (see equation (3.12) and Table 3.1), $\sigma^2_{\epsilon'_{cr}}$, (ii) the square of the total trade of each crop averaged over the period 1986-2011, $overline{W_{cr}}^2$, and (iii) the square of the global average crop water footprint over the same period, $\overline{CWF}^2_{cr}$, i.e.,

$$\sigma^2_{VWT} = \sum_{cr=1}^{cr=4} \sigma^2_{\epsilon'_{cr}} \cdot \overline{W}^2_{cr} \cdot \overline{CWF}^2_{cr} = \sum_{cr=1}^{cr=4} \sigma^2_{\epsilon'_{cr}} \cdot \overline{VW}^2_{cr}. \tag{3.13}$$

The product between $\overline{W}_{cr}$ (expressed in ton) and $\overline{CWF}_{cr}$ (expressed in m$^3$·ton$^{-1}$) gives the average water volume virtually embedded in the traded crops, i.e., $\overline{VW}_{cr}$.
The width of the 90% confidence interval with respect to the distance between the $CWF_{c,t}(Y)$ line and the $\overline{CWF}_{c,T}$ line suggests that assuming constant evapotranspiration over time has less impact on the $VWT$ estimates than the adoption of a time-constant crop water footprint.

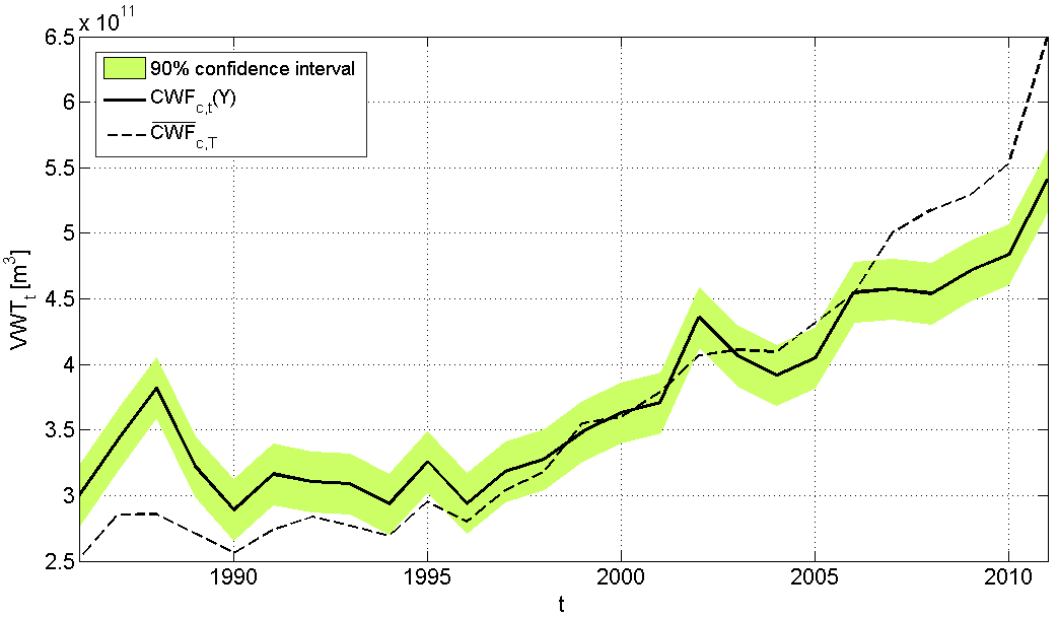


Fig. 3.5 Temporal trend of the virtual water volume associated to the international trade of wheat, rice, maize, and soybean in the period between 1986 and 2011. The black dashed line represents the virtual water trade evaluated with the time-averaged $\overline{CWF}_{c,T}$ values; the black solid line refers to the $VWT$ obtained using the annual $CWF_{c,t}(Y)$ values estimated with the FT method. The green area displays the 90% confidence interval of the FT method error, due to the assumption of constant crop evapotranspiration.

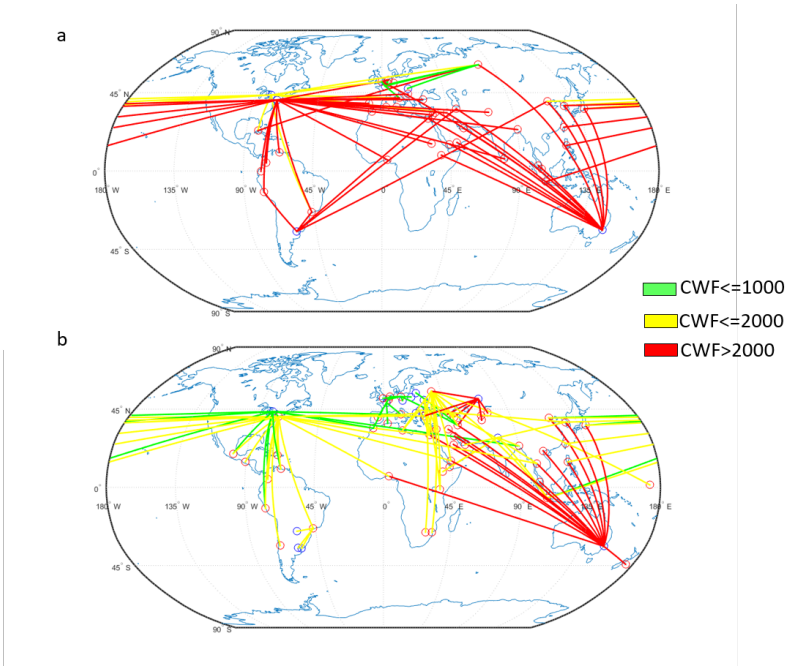


Fig. 3.6 The efficiency of virtual water trade in 1986 (a) and 2013 (b). Link color stands for the $CWF$ of wheat in the country of export: green indicates $CWF \leq 1000$ $m^3 \cdot ton^{-1}$, yellow indicates a $CWF$ in the range (1000;2000] $m^3 \cdot ton^{-1}$, red indicates that the $CWF$ is higher than 2000 $m^3 \cdot ton^{-1}$. Countries are represented through nodes: blue circles stand for exporting countries, red circles signify importing countries.

Finally, Figure 3.6/ shows the efficiency of the wheat international trade in years 1986 and 2013. The link efficiency is equal to that of the exporting country (the blue circle). While in 1986 most links are identified by the red color suggesting a virtual water content higher than 2000 $m^3$ per traded tonne, in 2013 the trade results more water efficient with most links showing an average associated virtual water content lower than 2000 $m^3 \cdot ton^{-1}$ or even lower than 1000 $m^3 \cdot ton^{-1}$, such as for the links departing from the US. Thanks to the $CWF$ improvements in many producing and exporting countries, the global trade of wheat in 2013 is less water intensive than the trade in 1986; in fact, along this period the amount of traded tonnes increases of 84% while the amount of embedded virtual water increased of only 42%.

## 3.4 Crop water footprint over 1961-2013

The annual $CWF$ estimates obtained considering both yield and evapotranspiration patterns (Section 3.1.3) are here analysed. Globally, all the crops experience $CWF$ decreases (Figure 3.7) along the period 1961-2013, mainly as a consequence of yield increases. Maize is the most water efficient crop across

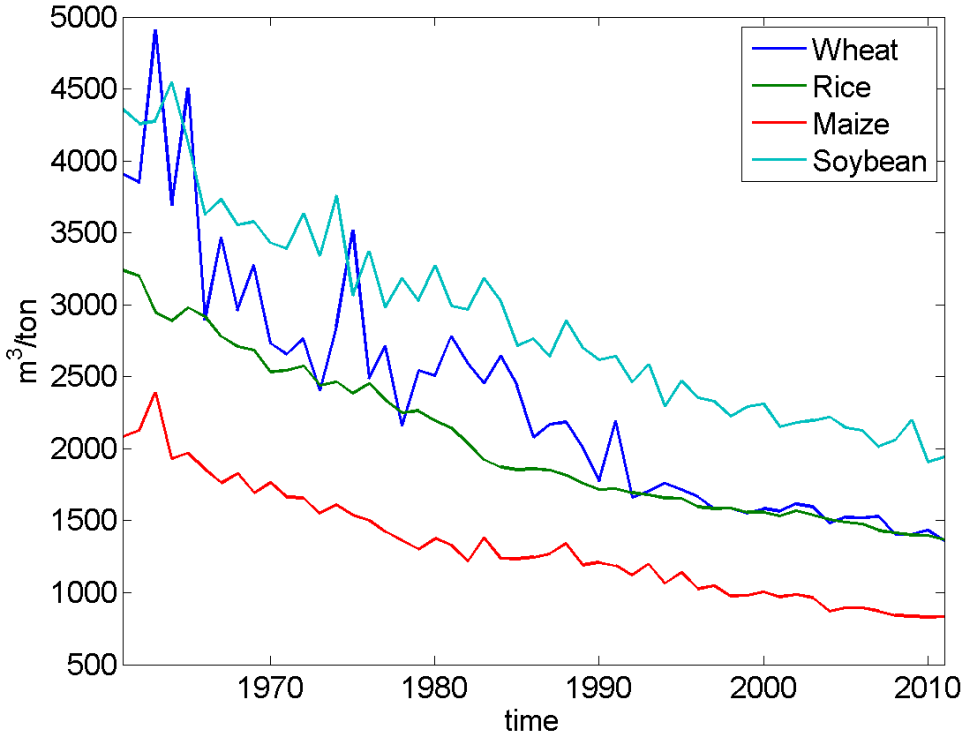


Fig. 3.7 Production-weighted global $CWF$ in the period 1961-2011 for wheat, rice, maize, and soybean.

the entire study period with a $CWF$ of 2000 $m^3 \cdot ton^{-1}$ in 1961 and around 800 $m^3 \cdot ton^{-1}$ in 2013. Maize global trend results regular with small fluctuations from year to year. Similarly, rice production is characterized by a regular negative trend of crop water footprint that overlaps that of wheat $CWF$ from year 1992. Before this year, wheat exhibits higher water footprint than those of rice and maize, and its trend results significantly variable with some $CWF$ peaks in 1963, 1965, 1974. In these years wheat water footprint overcomes the soybean water footprint, which is generally larger than all the other crops. Soybean production also shows a negative water footprint trend suggesting an

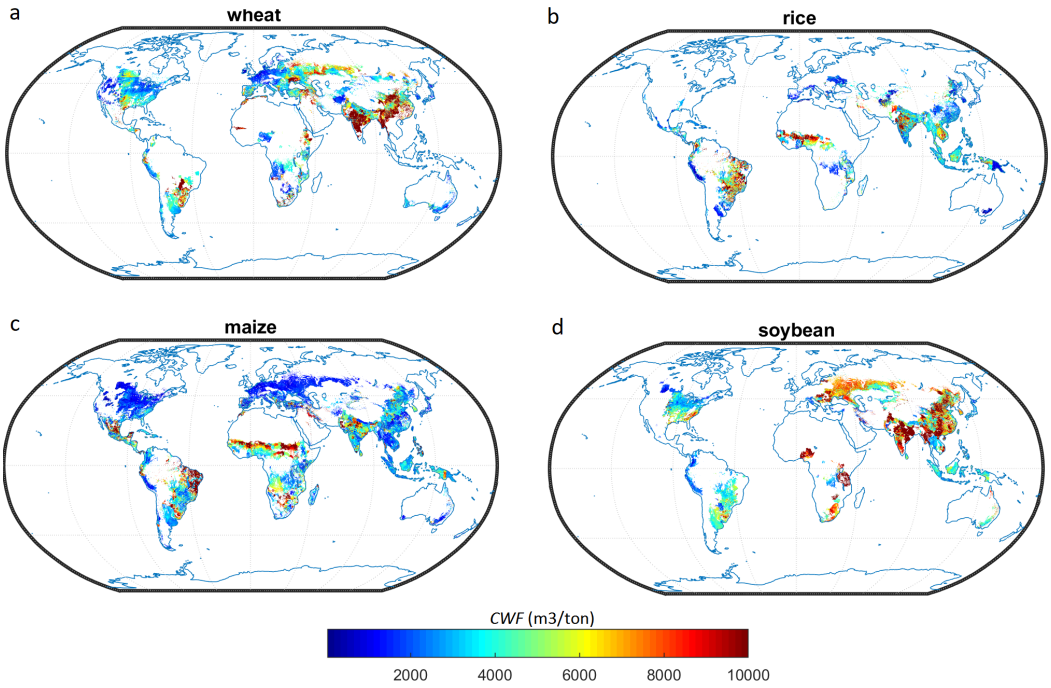


Fig. 3.8 Crop water footprint of wheat, rice, maize, and soybean in year 1961.

increasing water use efficiency. However, in 2013 soybean water footprint still remains more than twice that of maize.
Despite the global picture of water use efficiency improvements, the high-resolution maps in Figures 3.8, 3.9 suggest that there are areas where $CWF$ is still very high: e.g., in Africa, in the North East India for soybean production, in Central India for wheat production, and in North East Brazil for maize production. In particular, we find that wheat, maize, rice, and soybean crops experience water use efficiency improvements in 70, 90, 74, 83% of their harvested areas.

Wheat experienced $CWF$ decreases in around 70% of its harvested areas worldwide (Figure 3.10,a), with an average annual rate of -2,-0.1%/year. China and India show the largest annual rate of $CWF$ decreases (-2%/year and -1.2%/year, respectively), followed by Chile, Turkey, and Pakistan (rate of -1%/year). In the United States and Europe, $CWF$ decreases at lower rates, around -0.5%/year or it remains constant as in Colorado, New Mexico, Spain, and Russia. Africa is mostly dominated by constant or increasing crop water footprint (e.g., Morocco, Nigeria) with the exceptions of Ethiopia, Egypt,

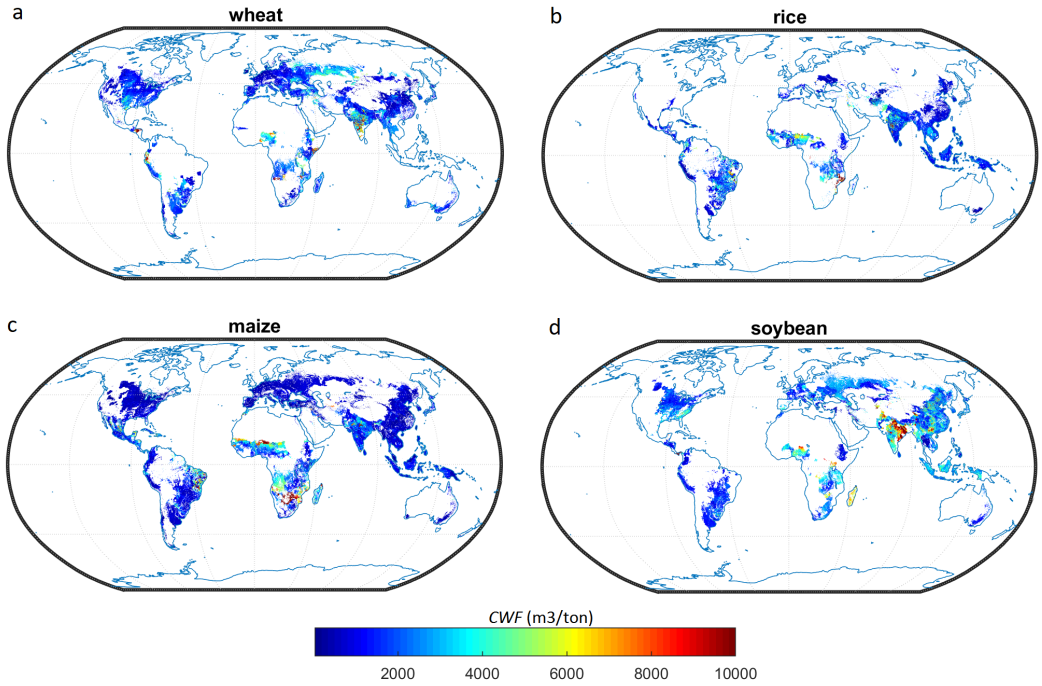


Fig. 3.9 Crop water footprint of wheat, rice, maize, and soybean in year 2013.

Mauritania, and South Africa where the $CWF$ decreases at a rate of 1%/year. Also Australia exhibits a constant crop water footprint.

Rice $CWF$ decreased in 74% of global areas: in large parts of Brazil, China, and Indonesia (Figure 3.10,b) with rates -1.5,-2%/year. $CWF$ decreases in Benin and Burkina Faso, but at lower rates (-0.3%/year), similar to those found in Argentina, Peru and Pakistan. West Europe does not experience $CWF$ decreases, while in East Europe some areas (e.g., Ukraine) improve the water use per unit production at a rate similar to that of India (-1%/year).
Around 90% of global maize-growing areas present crop water footprint decreases (Figure 3.10,c). In the Americas, maize $CWF$ decreases almost everywhere, but at different rates: -1.2%/year in Brazil and Mexico, 0.05%/year in the US and Argentina. The largest rate is found in Chile, i.e., -1.4%/year. Differently from wheat, African production of maize is mostly characterized by a constant or increasing crop water footprint, similarly to India and Eastern Europe. Spain, Turkey, and Greece show the fastest $CWF$ negative rates among the European countries (around -1.5%/year). Pakistan and China present average rates of -1.2%/year.

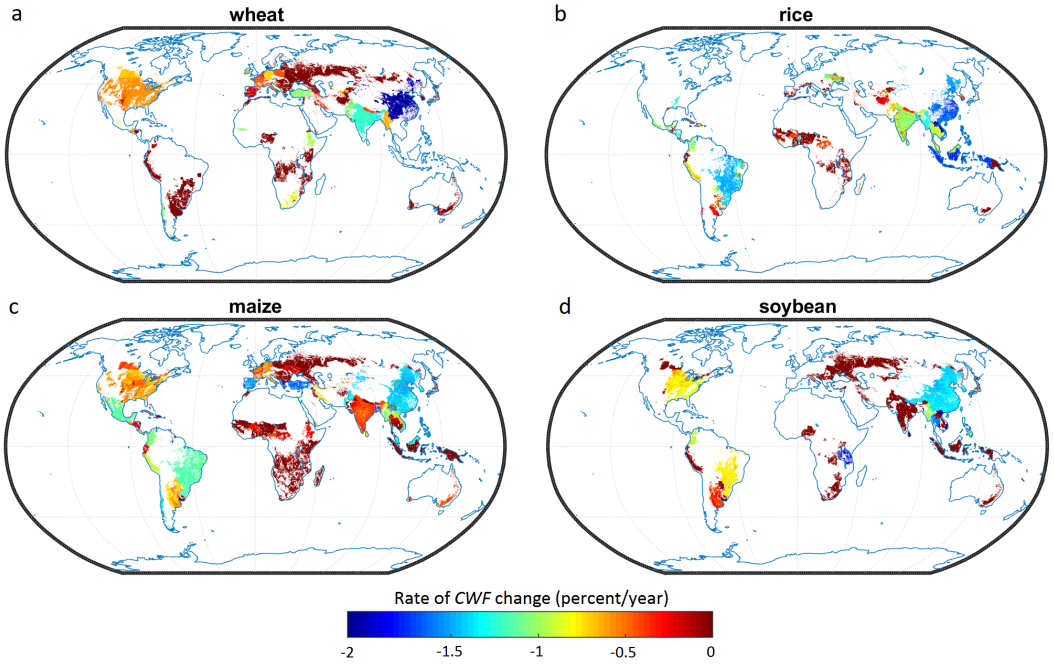


Fig. 3.10 Average annual rates of crop water footprint ($CWF$) changes for wheat (a), rice (b), maize (c) and soybean (d) production along the period 1961-2014. Negative values indicate a decrease in the water amount used to produce a unit crop; positive values indicate a constant or increasing $CWF$.

Soybean crop water footprint improved in 83% of global areas (Figure 3.10,d). Vietnam, China, and Thailand lead such increases with an average annual $CWF$ decreasing rate of -1.5%/year-2%/year. Tanzania as well exhibits a significant decrease in crop water footprint, which is even faster of that in the US and Brazil (around -1%/year), which are ones of the biggest soybean producers. India, Indonesia, and Australia do not present any improvements in the water use efficiency.

The United States, India, and China are the top-three producers of wheat. All these countries improved the water use efficiency along the study period as shown in Figure 3.10. However, they show different decreasing trends, both in terms of rates and relatively to the initial conditions of crop water footprint in 1961 (Figure 3.11). At the beginning of the study period, in fact, the United State is the most water efficient country among the top producers, with a $CWF$ of 3000 $m^3 \cdot ton^{-1}$. Conversely, China and India are less water efficient: i.e., $CWF$ of 5000 and 5700 $m^3 \cdot ton^{-1}$, respectively. From year 1972, China becomes more water efficient than the US: in just 12 years China reduced its $CWF$ of 3000 $m^3 \cdot ton^{-1}$. Then from 1973 to 1983, China further decreased its $CWF$ at a faster rate than the US; from 1984 the improvement continues, but at a lower rate. India improvement along the study period was as intense as that found for China, and it continues till the Nineties when the trend slowed down. From 1991 Indian water use efficiency is really similar to that of the US, with some years presenting also an inversion of the two countries (e.g., 1995, 2002). The US $CWF$ improvement is less marked than that of the other two producers, passing from an initial $CWF$ of 3000 $m^3 \cdot ton^{-1}$ to a $CWF$ of 1500 $m^3 \cdot ton^{-1}$.

The $CWF$ negative trends shown in Figure 3.11 are mostly explained by the yield positive trends (Figure 3.12). For all countries, yields show increasing trends over time, but the rate is different. China witnesses the fastest yield improvement: from 0.5 $ton \cdot ha^{-1}$ in 1961 to 5 $ton \cdot ha^{-1}$ in 2013, with an average annual rate of 0.09 $ton \cdot ha^{-1}$ per year. Wheat yield increased at 0.03 $ton \cdot ha^{-1}$ per year in the US and 0.05 $ton \cdot ha^{-1}$ per year in India. In year 2009 Indian yield overcomes the US yield.
Performing a Student $t$-test to determine the goodness of the linear model fit against the null hypothesis of a constant model, we found that the yield trends are significant for all the three countries with a significance level of 10%.

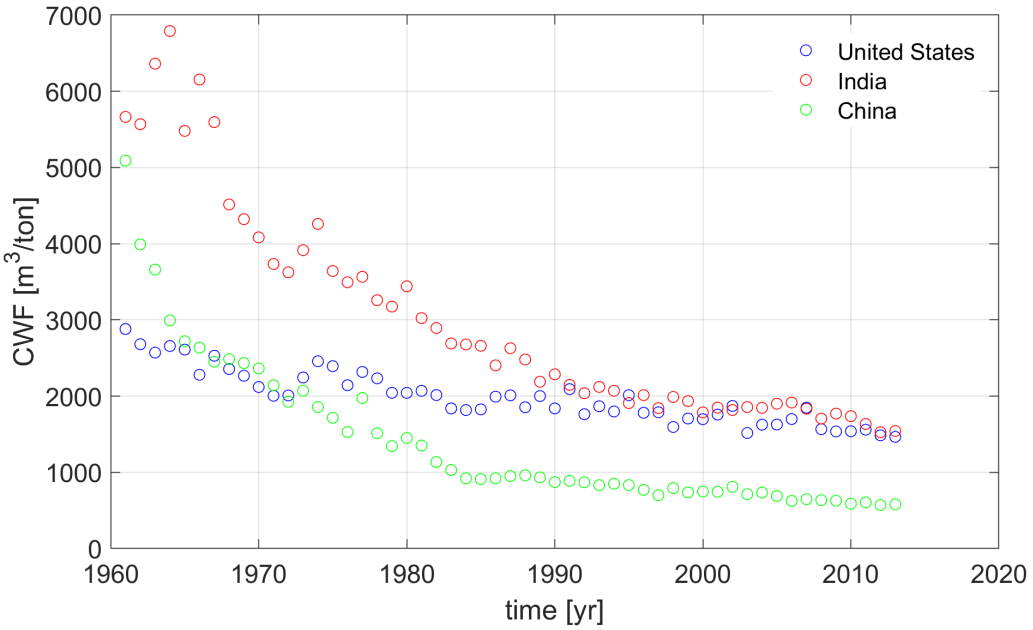


Fig. 3.11 Annual crop water footprint ($CWF$) of wheat in the United States, India, and China.

The evapotranspiration trend is significant in the US and China, while it is not significant in India. We found that in the US the crop evapotranspiration increases at only 0.33 mm/year in the US and 0.47 mm/year in China, though the trend significance.

Figure 3.14 shows the coefficient of variation of annual $CWF$ at the grid level. It expresses the extent of crop water footprint variability in relation to the time-averaged $CWF$. Indeed, in each cell we considered the $CWF$ time series from 1961 to 2014, and we calculated the ratio between the standard deviation and the mean.
The coefficient of variation is highly variable across countries and crops. Major variations are observed in Peru, Brazil, South Africa, India, and China for wheat production. Rice, maize, and soybean result more stable over time: in many countries the variation is lower than 25% of the mean (e.g., Latin America, India). However, Spain, some African countries and China show variation around 40% for maize production. Soybean $CWF$ is particularly unstable in South Africa, Eastern Europe, and Australia; conversely, in the US and Brazil the soybean water footprint shows a coefficient of variation of 10-20%.

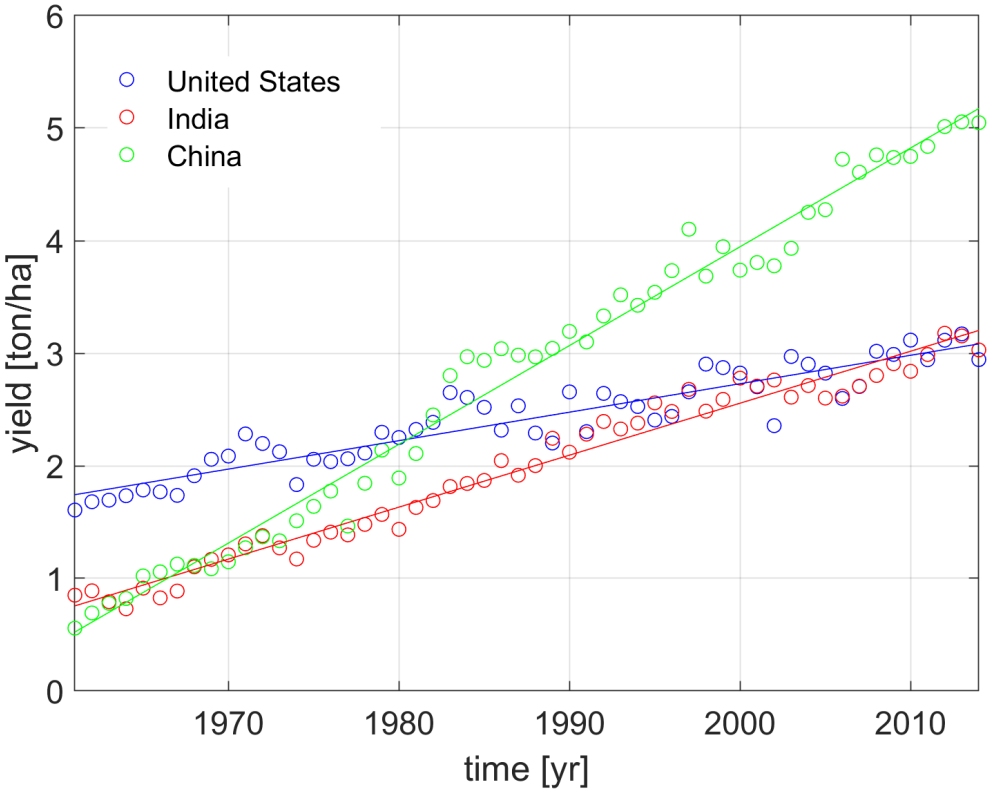


Fig. 3.12 Annual wheat yield in the United States, India, and China. The solid lines are the linear regression model fits to data and coloured according to the countries.

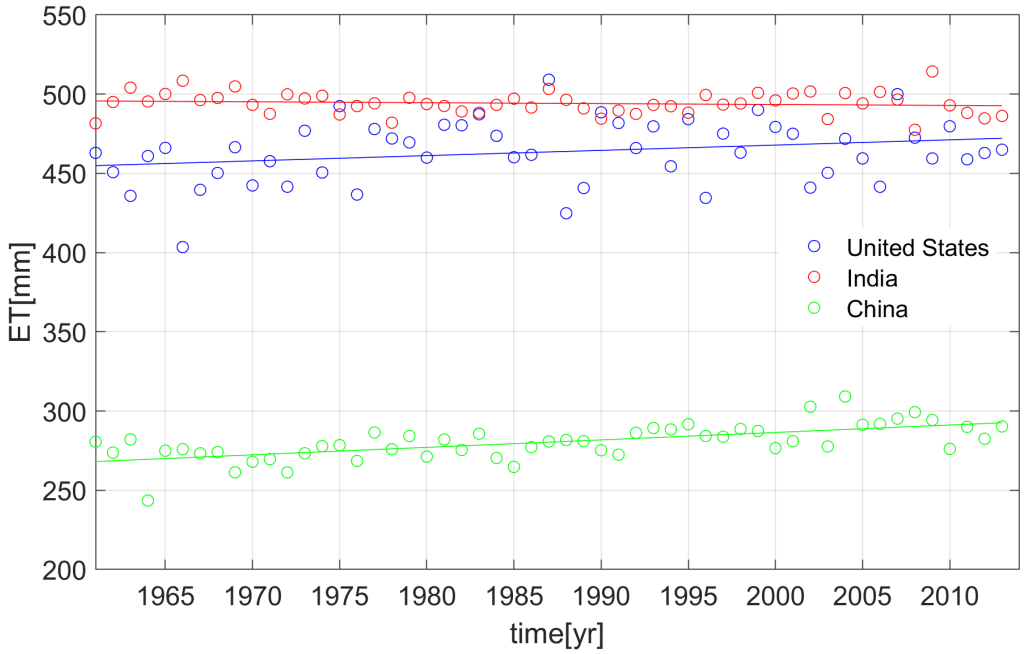


Fig. 3.13 Wheat evapotranspiration depth over the growing season in the United States, India, and China. The solid lines are the linear regression model fits to data and coloured according to the countries.

## 3.5 Concluding remarks

In this Chapter, we demonstrate the feasibility of the Fast-Track approach to provide estimates of time dependent crop water footprint. The method is tested by comparing the annual $CWF$ country-values of wheat, rice, maize, and soybean obtained through the FT approach with those obtained by a detailed model accounting for the changes of both yield and evapotranspiration over time. The two estimates compare well with a coefficient of determination close to 1 for all crops. This suggests that inter-annual variations of crop water footprint is mostly driven by yield variability, while the effects of evapotranspiration not embedded in yield variations [111, 110] seem to be marginal when compared to yield, thus confirming the assumption of the FT approach.
To accomplish the assessment of the FT approach, the uncertainty of the FT method, due to considering time-constant evapotranspiration, has been assessed finding a general low uncertainty of the $CWF$ estimates with a standard deviation of the error around 0.1. Such uncertainty is three-times lower than that of the model used to estimate the crop water footprint. Finally, the

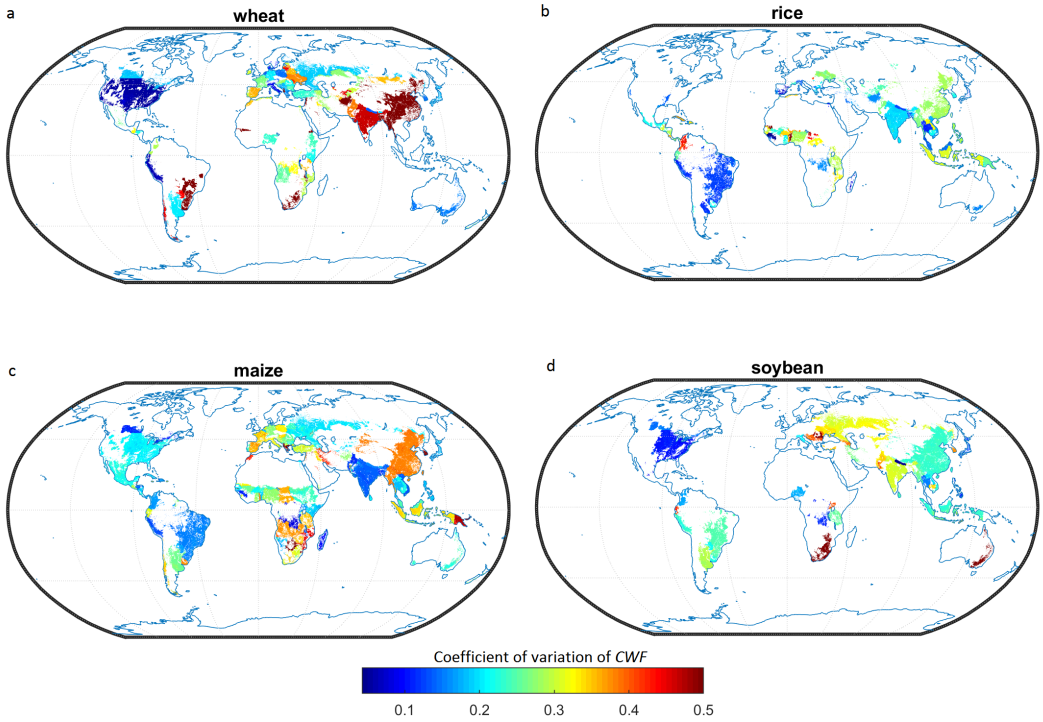


Fig. 3.14 Coefficient of variation of the annual crop water footprint over the period 1961-2014. The ratio of the standard deviation over the entire study period to the average $CWF$ over the same period.

time-dependent crop water footprint estimates have been applied to evaluate the virtual water volume associated to the international trade of wheat, rice, maize, and soybean over the period 1986-2011. Comparing this pattern with the one obtained using constant $CWF$ values, as previous studies did [22],[132], confirms the importance of including time-dependent crop water footprint in the computation of virtual water trade. The results prove the suitability of the FT approach, which represents a very useful tool thanks to its low computational cost, and its easy and fast applicability.

# Chapter 4

# The Water Debt repayment time

Agriculture vastly dominates global freshwater consumption, drawing water resources from soils, surface and underground water sources, which are naturally renewed at different rates. As above mentioned, water resources are dangerously stressed and overexploited due to human water consumption. Particularly, in some regions people are living in highly water-stressed areas [83], and two thirds of the global population lives under severe water stress conditions for at least one month a year [93]. The General Assembly of the United Nations has set seventeen Sustainable Development Goals (SDGs) to stimulate action to protect the planet toward year 2030 [11]. Specifically, SDG number 6 (target 6.4) aims at ensuring sustainable use of water resources in order to reduce the number of people suffering from water scarcity [69, 141]. In this Chapter, a quantitative answer to this urgent matter of promoting sustainable water use is proposed through a new metric, which we call "water debt repayment time". This indicator builds upon a broad context of well-known water shortage and water scarcity metrics, and aims at providing a physical quantification of water use sustainability, measured by the time required to replenish the water resource used, splitting the role of soil, surface, and groundwater. This indicator also assess major responsibilities behind the overuse of water resources in agriculture, through a crop-specific and spatially-explicit analysis.

In the last decades, a large number of indicators has been introduced to monitor the (mis)match between water demand and availability [63, 147, 144, 141].

These studies underpinned regions with major water insecurity through analyses at country [120] or grid level [93, 127], on annual [144] or monthly basis [74, 117], also integrating the two temporal scales [19] and providing outlooks to possible future scenarios [144]. Most water scarcity metrics, or criticality ratios, are based on risk categories, such as "high water stress" if the use-to-availability ratio is higher than 0.4 and "very high water stress" if greater than 0.8 [4]. Risk categories metrics lack impartiality because the critical level of stress is arbitrary and is not coherent across studies.
In the framework of life cycle assessment, other indicators have been introduced to assess the potential environmental impact of water use [82]. These indicators quantify the potential impact based on freshwater use inventory schemes [17], weighted by local characterization factors which transform inventory flows into environmentally equivalent flows [104]. All these indicators have proven useful to assess the geographic and temporal mismatch between water demand and availability under different perspectives. However, each indicator has some shortcomings. First, water scarcity metrics generally focus only on blue water resources, without considering the interplay between blue and green water scarcity [141]. Moreover, only in few recent studies blue water use has been splitted into surface -and ground-water use [144]. Second, these indicators mostly lack a physical interpretation, being generally based on risk categories or potential impact factors. Third, the causes of the scarcity are rarely traced back to their specific determinants, i.e., the particular crop generating the mismatch between water use and availability. The study by Dalin et al. [33] is the first to analyse crop-specific responsibilities behind groundwater depletion embedded in international food trade, but it did not consider surface water use or green water use. The concept of water debt repayment time addresses all the above mentioned issues, and summarizes, in a single metric all the advances introduced by recent studies. This indicator takes into account green, surface and ground water resources and enables source-specific analyses across crops and locations.

## 4.1 Water Debt repayment time: definition

Water use for annual crop production is unsustainable when it exceeds the amount of water annually available from the local water cycle. When local availability is exceeded, a certain time is required to replenish the water resource that has been used to achieve the annual crop production. We call this the "water debt repayment time". The water debt repayment time ($WD$) is calculated as the ratio of the source-specific water footprint in a grid cell (5 arc minute resolution), to the amount of water annually available in the same cell. In particular, this indicator measures the time required to replenish the water source $s$ used by crop $cr$ in cell $l$, i.e. $WD_{s,cr,l}$, and results from the ratio of the water depth evapotranspired in the cell during all growing seasons in a year (i.e., the cell water footprint, $WF_{s,cr,l}$ [$m^3$]) and the annual renewability rate of the same source in cell $l$, $R_{s,l}$ [$m \cdot yr^{-1}$] (see Section 4.3), multiplied by the cell area, $A_l$ [$m^2$], that is,

$$WD_{s,cr,l} = \frac{WF_{s,cr,l}}{A_l \cdot R_{s,l}} \qquad [yr]. \tag{4.1}$$

$WF_{s,cr,l}$ can be obtained as the product of local crop water footprint by source and the crop production ($PR_{c,l}$) in tonnes, that is

$$WF_{s,cr,l} = CWF_{s,cr,l} \cdot PR_{cr,l}. \tag{4.2}$$

When the water used for crop production is lower than (or equal to) the renewability rate, the $WD$ is recovered within the hydrological year, i.e. $WD \leq 1$ yr and, thus, the water use is sustainable; otherwise, the water depletion is repaid by next hydrological cycles, i.e. $WD > 1$ yr, indicating an unsustainable water use. The $WD$ of source $s$, arising from all cultivated crops at location $l$, $WD_{s,l}$, equals the sum of debts generated by each crop and reads

$$WD_{s,l} = \sum_{cr=1}^{cr=9} WD_{s,cr,l}. \tag{4.3}$$

Owing to the simultaneous replenishment of soil, surface- and ground-water by precipitation, the total $WD$ across the threes sources, $WD_{cr,l}$, is given by the

maximum $WD$ value obtained with equation 4.1), i.e.,

$$WD_{cr,l} = \max(WD_{sm,cr,l}, WD_{sw,cr,l}, WD_{gw,cr,l}) \tag{4.4}$$

where $sm$, $sw$ and $gw$ indicate the soil moisture, surface water and groundwater, respectively. We took the maximum water debt instead of the sum of the three water debts because every year precipitation recharge all the sources at the same time.
The water debt in a given area, being it a river basin, a country or the whole world, referred to a single crop $cr$ and source $s$ is evaluated as a production-weighted mean, i.e.

$$WD_{s,cr,L} = \frac{\sum_{l \in L} WD_{s,cr,l} \cdot PR_{cr,l}}{\sum_{l \in L} PR_{cr,l}}, \tag{4.5}$$

in order to highlight hotspots of large $WD$ generated by large production. $L$ is the ensemble of cells in the area of interest. In turn, if all crops are considered together, the $WD$ in the area (equation 4.3) is weighted by the water volume used by all crops in the cell (or cell water footprint), i.e.

$$WD_{s,L} = \sum_{l \in L} \Big( WD_{s,l} \cdot \frac{WF_{s,l}}{\sum_{l \in L} WF_{s,l}} \Big), \tag{4.6}$$

given the unfeasibility of production sum across crops.
These averages are considered instead of the simpler ratio of cumulated $WF$ and cumulated availability over the area $L$ because we account for spatial heterogeneity and acknowledge that consumption and availability of water may occur in different portion of the area, without the possibility of redistributing the resource.

## 4.2 Green and blue crop water use

The green and blue water footprint computed in Chapter 2 are here extended to barley, potatoes, sugar cane, sugar beet, and cotton. All the nine crops together account for over 70% of the global caloric content and 66% of the proteins in global food consumption [34].
Crop water use can originate from soil moisture, i.e. green water, and surface-

or ground-water bodies, i.e. blue water. Rain-fed and irrigated areas are available from the MIRCA dataset [108] and they are specific for each study crop. While usually the two sources of blue water are computed together [5], here we partition the two contributions proportionally to the areas equipped for irrigation with surface water ($AEI_{sw,l}$) and groundwater ($AEI_{gw,l}$). We thus assume that the ratio of $CWF_{gw,cr,l}$ ($CWF_{sw,cr,l}$) to total blue water footprint ($CWF_{b,cr,l}$) is equal to the ratio of $AEI_{gw,l}$ ($AEI_{sw,l}$) to $AEI_l$, i.e.

$$CWF_{s,cr,l} = \frac{AEI_{s,l}}{AEI_l} \cdot CWF_{b,cr,l}, \tag{4.7}$$

with $s = sw, gw$. This assumption may bring uncertainty in the estimation of surface- and ground-water uses if actual and potential use of irrigation in equipped areas differ, or if the ratio of ground- to surface-water irrigation varies across crop plots or seasons in the same area [125]. However, multiple studies have used the $AEI$ ratio to compute the irrigation crop water use per water source [125, 37, 35] although no differentiation among crops has been proposed. Results of groundwater use for irrigation purposes at the country scale obtained in this study have been compared with the estimates provided by *Wada et al.* [145] for total crop production (Figure 4.1). Each circle represents a country and the size is proportional to the importance of the irrigation sector in each country with respect to the other blue water uses (e.g., manufacturing and municipal uses). Such importance is evaluated as the ratio between the agricultural water withdrawals and the total water withdrawals derived from the AQUASTAT database (http://www.fao.org/nr/water/aquastat/data/query/index.html?lang=en). Hence, the smaller circles are for countries where irrigation water is less important; the bigger circles, instead, are for those countries where the irrigation sector is more important. Smaller circles are generally over the bisector given that the groundwater volume computed in this study only refers to the 9 crops production. Conversely, larger circles are closer to the bisector line. The circle color is intended to add another information about the importance of the nine crops irrigated area in respect to the total irrigated area. The darker is the circle the most important is the 9 crops groundwater demand with respect to that of other crops. The estimates from the two studies compare well, especially for the major groundwater-consuming countries (e.g., India, the US, Pakistan).

It should be noted that indications of actual irrigation volume at the global scale are generally lacking, but if local data were available, they could be used to derive a specific and more precise measure of surface- and ground-water use.

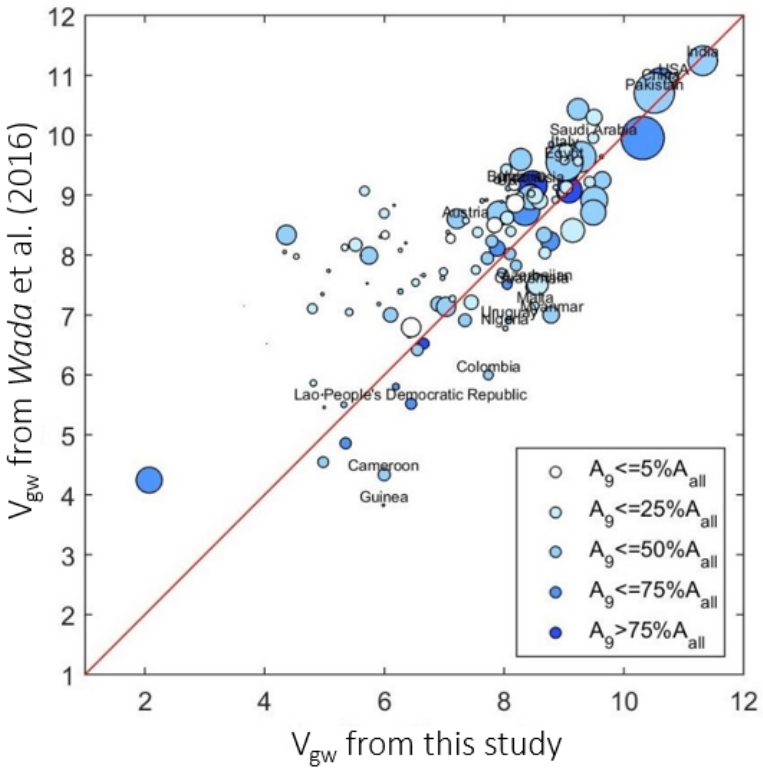


Fig. 4.1 Country scale comparison of the groundwater volume computed in this study for the 9 study crops with the groundwater volume obtained by [145] for all water uses (i.e., irrigation, manufacturing, and municipal uses). Circle size is proportional to the importance of the irrigation sector in each country with respect to other uses. Circle color represents the fraction of the irrigated area of the nine crops in respect to the total irrigated area.

## 4.3 Renewability of water by source

Renewable freshwater is the water flow generated from precipitation that is available to meet human and ecosystem needs [77]. In this study, it is considered in the form of soil moisture, surface water, and aquifers and data have been

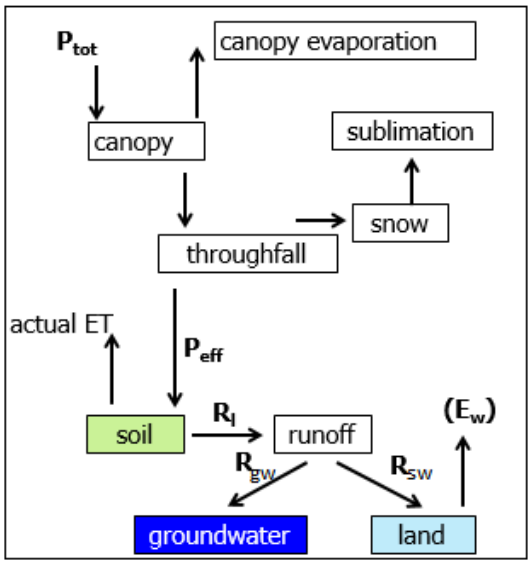


Fig. 4.2 Schematic representation of the major water fluxes (arrows) and storages (boxes). $P_{tot}$ is the total precipitation that reaches the cell, $ET$ is the actual evapotranspiration of the cell, $P_{eff}$ is the effective precipitation, $R_l$ is the total runoff, $R_{gw}$ is the portion of runoff that recharges the groundwater, and $R_{sw}$ is the surface runoff. The figure has been adapted from [96].

derived from the WaterGAP 2.2b dataset [96] obtained in a "no use" setting to simulate the natural recharge rates of each source. Figure 4.2, which has been adapted from [96], shows a schematic representation of the major water fluxes and storages considered in this study.

**Renewable soil moisture** is the fraction of effective precipitation, net of surface runoff and groundwater recharge ($R_l$ in Figure 4.2), that infiltrates into the upper soil layer and recharges the soil water storage, becoming available for root water uptake and evapotranspiration.

**Renewable surface water** is the net surface runoff ($R_{sw}$), produced from precipitation at a certain location, that flows to surface water bodies minus the evaporation from lakes and wetlands. It is negative when evaporation is larger than runoff, which happens occasionally (3% of the cultivated cells worldwide) in the dry regions of Egypt, Botswana, and Malawi, where precipitation is lower than the overall evapotranspiration/evaporation losses from land and water bodies. Negative monthly values of net runoff are set to zero in order to avoid computational problems in the $WD$ evaluation.

**Renewable groundwater** is the recharge ($R_{gw}$) originated from precipitation that deeply percolates the soil layers and reaches the aquifer. The recharge

can be both local (coming from surface water) and diffuse (coming from the unsaturated soil) [36]. The renewability rates of each water source are taken as the annual volumes of renewable water. Data from the WaterGAP 2.2b dataset are given as gridded data at 30'x30' spatial resolution, then values (in $m^3/m^2$) are considered uniform over each cell and downscaled to the finer (5'x5') grid. Monthly values from 1987 to 2013 (centered around year 2000) have been used to construct long-term average monthly values, then cumulated over the year to define the annual renewability rates.

## 4.4 Results

### 4.4.1 Global assessment of water use sustainability

Over the 1996-2005 period, the production of the nine crops required annually 3313 $km^3 \cdot yr^{-1}$ of water, with 82% coming from soil moisture and 18% from surface-(60%) and ground-(40%) water (Figure 4.3(b)). The commodities analyzed here exhibit different levels of water-related sustainability (Figure 4.3(c)). According to the global averages, maize, soybean, barley, potatoes, and sugar crops (providing 32% of the global food calories [34]) are water sustainable, while wheat and rice (providing another 36% of the global food calories [34]) are water unsustainable (Figure 4.3a). On a global average, cotton, rice, and wheat show $WD$ with groundwater larger than 1 yr, pointing out an unsustainable groundwater use. This indicates that these crops are irrigated with relatively slowly-replenished aquifers, particularly cotton ($WD$ of 4 yr). Cotton is also depleting surface water resources, with an associated $WD$ of 2 yr although it consumes about 30 $km^3$ of water globally (Figure 4.3b), which is much lower than the water consumption of wheat (56 $km^3$) and rice (196 $km^3$) that show a sustainable use of surface water on a global average. The water debt, $WD$, mainly depends on the water use efficiency ($CWF$, Figure 4.3b), on the volume of water used from different sources ($WF$, Figure 4.3c), and the local water renewability rate. The efficiency of water use varies significantly across the nine crops (Figure 4.3c). Cotton is by far the most water-intensive crop (nearly 4000 $m^3 \cdot ton^{-1}$), but sums to a relatively small global water use (192 $km^3$), because it is not grown as much as less water-intensive crops such as rice (1435

$m^3 \cdot ton^{-1}$, 790 $km^3$ total), wheat (1529 $m^3 \cdot ton^{-1}$, 845 $km^3$ total) and soybean (2243 $m^3 \cdot ton^{-1}$, 368 $km^3$ total). Over seventy percent of the irrigation water used to grow the nine crops is used to produce rice (289 $km^3$) and wheat (120 $km^3$). Barley, potatoes and sugar crops are all relatively less water-intensive and lead to a relatively lower global water use (Figure 4.3c), making their production sustainable in terms of water use (Figure 4.3a). The global-scale values of water use and water debt give a first insight into the water sustainability of different crops. However water debts are heterogeneous across and within countries, underpinning areas of higher or lower sustainability related to the use of different water resources (Figure 4.6).

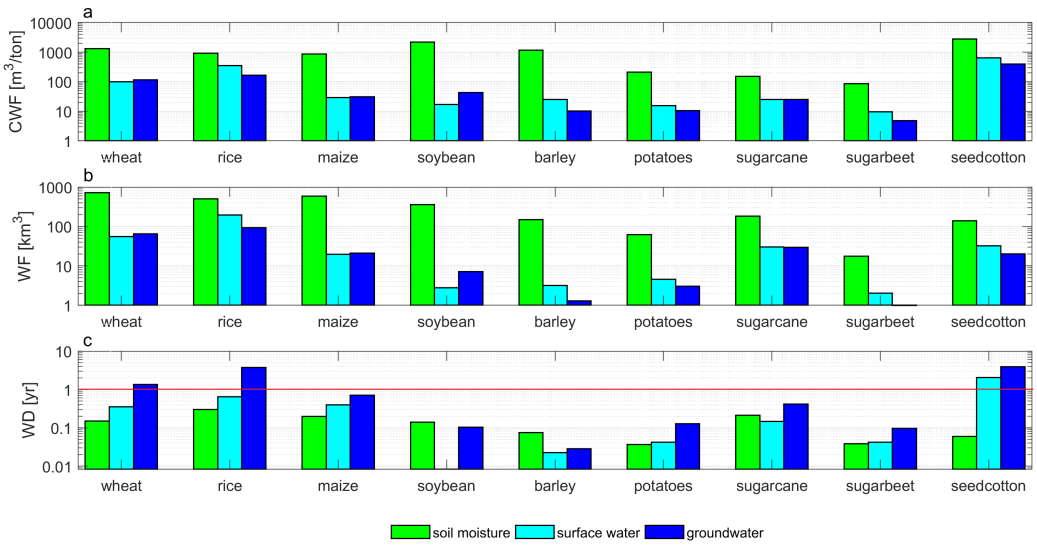


Fig. 4.3 Globally-averaged crop water footprint ($CWF$, a), global volume of water use ($WF$, b), and globally-averaged water debt ($WD$, c) for each of the nine crops and three water sources. Note that the y-axis is shown in log-scale.

### 4.4.2 The water footprint of crop production

Figure 4.4 shows the total green water footprint of crop production at 5'x5' arc minute resolution. This map underpins locations where water use from soil moisture is higher. The highest green water volume are found in the US core production where $WF_g$ is higher than 25 million $m^3$ in many cells; maize and soybean production exploit the biggest portion of available green water during the year. Also Brazil and Argentina show high green water use especially for

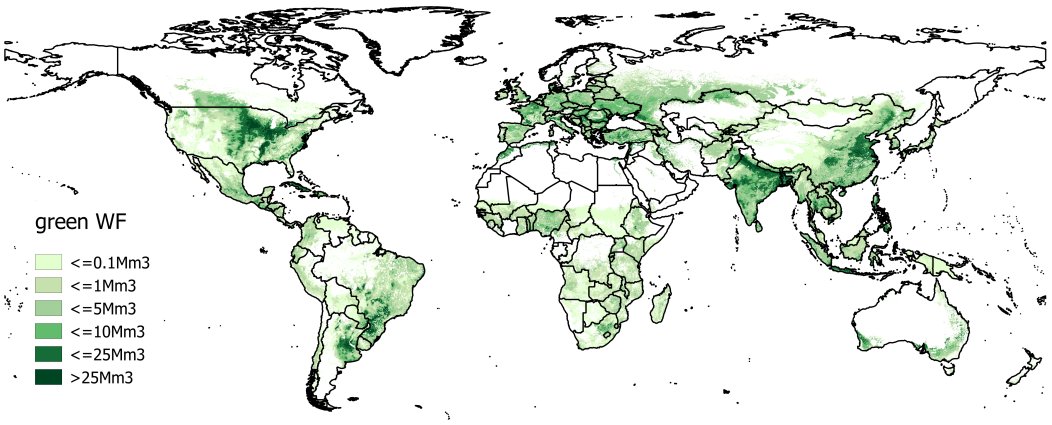


Fig. 4.4 Map of green water footprint of crop production in year 2000.

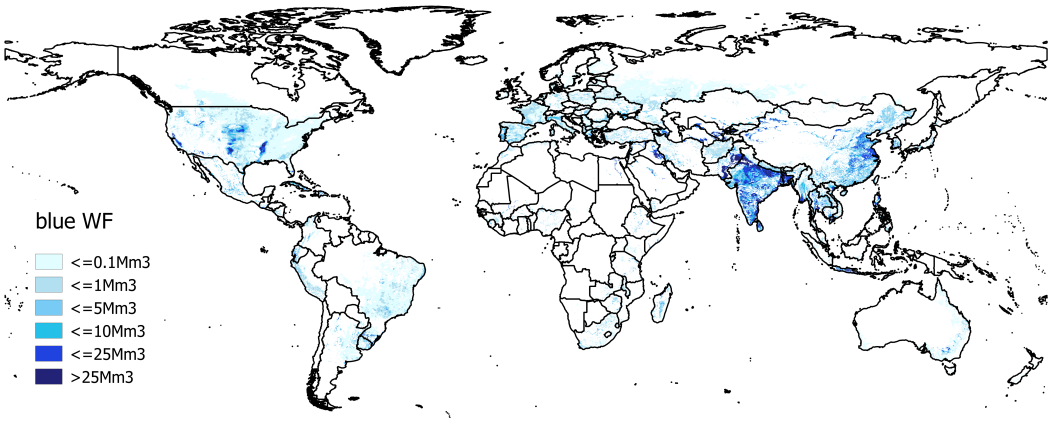


Fig. 4.5 Map of blue water footprint of crop production in year 2000.

the soybean and sugar cane production. Large green water footprints are also found in India and China for wheat, maize, and rice production.

Figure 4.5 depicts the total blue (surface plus ground) water footprint of crop production. Similarly to $WF_g$, $WF_b$ is unevenly distributed in space, with larger water use closer to the most important aquifers (e.g., the California Central Valley and the High Plain aquifer in the US) and major river basins (e.g., the Mississippi river basin and the Indus River basin). Large blue water footprints are also found along the river courses: e.g., the Amu Darya river that has been transformed into an irrigation channel to watering cotton production in Uzbekistan, the Yellow river which mostly sustains rice and maize production, the Tarim river that feeds rice.

## 4.5 Geography of the water debt repayment time

Figure 4.6 shows the water debts with soil moisture, surface, and ground water bodies associated to the production of the nine study crops.

**The sustainability of green water use** The water debt arisen by green water use, $WD_{sm}$, measures the number/fraction of years during which water is not available to satisfy other competing demands for soil moisture, e.g., watering other crops, grazing lands, forestry, or to sustain the terrestrial ecosystems. $WD_{sm}$ associated with the production of the nine study crops, is always lower than one year (Figure 4.6a), denoting a fully sustainable rain-fed production; particularly, 60% of the cultivated cells show $WD_{sm}$ values lower than 0.25 years. Nonetheless, we found cells where $WD_{sm}$ is close to one year, mainly due to an intense agricultural production. The largest green water exploitation is located in the US core production area (black circle, Figure 4.3(a)) where nearly 40% of the (nine crops) national production is located. Here precipitation is the most important water source for crops (mostly, for soybean and maize), making it not available for other uses for 6-8 months per year. Similar conditions are found in some areas of France, Germany, and Italy, where soil moisture exploitation is more spread out across the country. Northern India presents $WD_{sm}$ higher than 8 months due to wheat, sugarcane, and rice production, while Eastern China exhibits $WD_{sm}$ higher than 6 months mostly because of maize production.

**The sustainability of surface water use** Annual crop production incurs water debt with surface water resources when irrigation demand for surface water irrigation exceeds the locally generated runoff. Since we focus on the local sustainability of water use, we do not consider upstream flows as available sources in downstream cells. This assumption may lead to overestimates of $WD$ in downstream cells, but it allows to clearly underpin all the areas that are not locally sustainable because of (i) the overexploitation of local water resources, (ii) the low renewability rate of local water resources, or (iii) the dependence on upstream water resources, which is transboundary in some cases

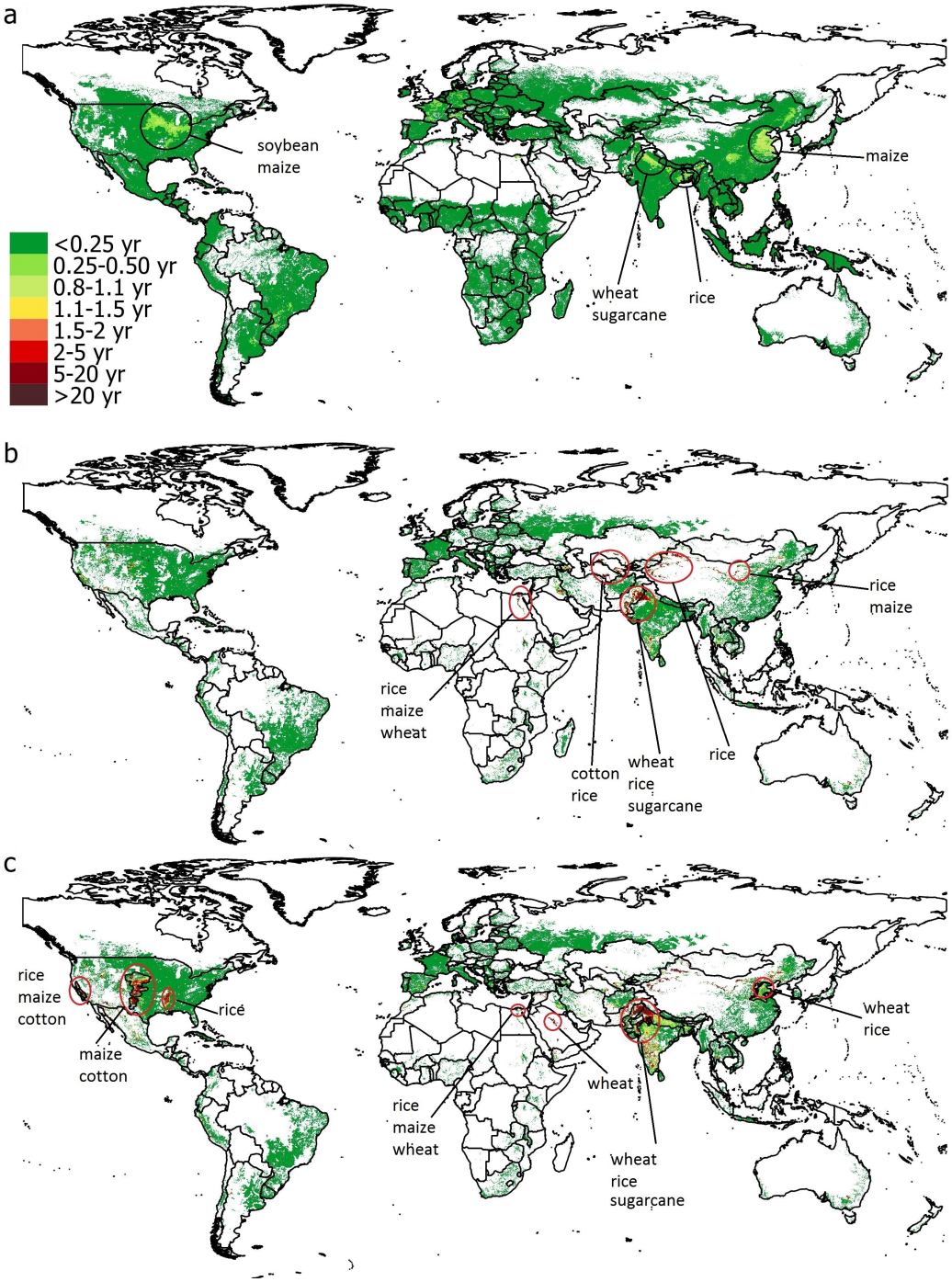


Fig. 4.6 Water Debt ($WD$) associated with the production of nine major crops. The $WD$ value in each cell (5'x5' spatial resolution) is expressed as the number of years required to replenish the water source used for crop production: soil moisture (a), surface water (b), and groundwater (c).

(e.g., Nile and Rio Grande [97]). In order to stress the different levels of water use sustainability, we calculate the surface water footprint and the $WD$ both at the grid and basin level. These two different outputs could highlight the importance of scale in water resources management. In fact, the surface $WD$ at the basin scale is calculated as the basin' surface water footprint divided by the total runoff generated within the basin, which may be provided by water infrastructures.

At the basin level, all basins sustainably feed the production of the nine crops examined as shown in Table 1; though, sustainability issues may arise due to the water used by other crops/sectors and if considering environmental flows requirements. The basins of Sabarmati in India and Chao Phraya in Thailand show the least sustainable surface water footprint, showing longer $WD$ than to all the other basins: i.e., 0.33 and 0.22 year (Table 1, or 120 and 79 days, respectively). The surface water footprints of these basins are 24 and 7.8 $km^3$, respectively. The surface water footprint of the Sabarmati basin is smaller than that of Indus and Yangze basins, but a longer $WD$ is required because of the lower surface water availability. Sugarcane and rice are the crops mostly responsible for the large repayment time there. Locally, the replenishment time is longer than 1 year in the North West and South of the basin, but $WD$ values are always shorter than 8 years. Conversely, longer $WD$s are found across the Nile Delta, where $WD$ reaches 20 years, due to the cultivation of rice, maize, and wheat (Figure 2B). Here, irrigation demand is mainly reliant upon upstream cells as confirm by the study by Munia et al. 2017 [97] and by the low $WD$ of the whole basin. Crop production spread over the Indus River Basin generates locally unsustainable water use, with $WD$ longer than 20 years in some cells in the East of the basin. We found that wheat, rice, and sugarcane production draws most of the water annually available from locally generated runoff. The local reduction of river flow due to withdrawals and the low renewability rates are responsible for such large WD values. However, when we consider the water debt at the basin scale we obtain an average $WD$ of 0.15 year (or 53 days, Table S1). Other vulnerable areas are found in China, along the Tarim River, where the largest $WD$s are mainly due to rice production, and along the Yellow River, where rice and maize are most responsible for the overuse of surface water. Finally, due to the diversions of the Amu Darya and Syr Daria rivers to grow cotton and rice in an arid region, the areas close to

the Aral Sea have undergone serious environmental damages [103] and show long $WD$.

Table 4.1 Surface water footprint and water debt of major river basins, which sustain 50% of the global surface water footprint due to the cultivation of the nine crops. Basin delimitation is provided by the GRDC repository [1].

| Basin name | Surface WF [km$^3$] | Surface WD [yr] |
|---|---|---|
| Sabarmati | 24.3 | 0.33 |
| Chao Phraya | 7.8 | 0.22 |
| Caspian Sea, East Coast | 6.1 | 0.16 |
| Indus | 36 | 0.15 |
| Tarim Interior | 6.5 | 0.14 |
| Krishna | 9.8 | 0.12 |
| Amu Darya | 9.8 | 0.10 |
| Syr Darya | 7.3 | 0.08 |
| Java-Timor | 12.3 | 0.08 |
| China Coast | 23.5 | 0.08 |
| Gobi Interior | 5.2 | 0.08 |
| Huang He | 4.9 | 0.06 |
| Nile | 19.3 | 0.06 |

**The sustainability of groundwater use** The water debt with groundwater, $WD_{gw}$, occurs when groundwater use by crops exceeds ground water recharge (i.e., deep percolation, not including irrigation return flow). Over half of the groundwater used worldwide for the nine crops originates from just four major aquifers, namely the Indo-Gangetic plain (41%), U.S. High Plain (8%), North China plain (5%), and the California Central Valley (1.6%) aquifers. The highest groundwater use is found in the Indo-Gangetic plain (100 km$^3$/yr), where 64% of the Indian and Pakistan crop production is located. The average $WD_{gw}$ over the I-G plain aquifer is around 13 years, but some zones also reach $WD_{gw}$ of 100 years. This means that water use for irrigation is markedly depleting the aquifer and the time required to replenish it extends to future generations. In particular, the highest $WD_{gw}$ are found in the Upper Ganges, while the Lower Ganges appears to be sustainably exploited, due to lower water use and higher recharge rates [59]. Transferring the production in Lower Ganges might be a possible solution to lower the water debts, but not to achieve full sustainability. In order to improve the sustainability of crops produced

in these areas, water withdrawals could be reduced by improving water use efficiency. Over the U.S. High Plain area, the groundwater exploitation is significant [88] and the average $WD_{gw}$ is around 7.2 years, but even higher $WD_{gw}$ values are found in the central and southern part of the HP (i.e., Kansas and Texas mostly), where the renewability rate is lower than that in the North [116]. Among the study crops, maize and cotton are the main responsible for the groundwater depletion, as *Dalin et al.* 2017 [33] also pointed out. Over the California Central Valley water debt increases form North (2 yr, due to rice and maize production) to South (9 yr, due to cotton production). Groundwater-fed crop production in the North China Plain appears to be unsustainable as well, but the $WD$ caused by the crops production (mainly wheat and rice) is smaller than that found over other aquifers, i.e. $WD_{gw}$=1.5 yr. Northern China shows higher $WD_{gw}$, mostly associated to rice production (e.g., Xinjiang province). Our estimates of unsustainable use of groundwater resources are in accordance with previous studies, e.g. the groundwater footprint indicator developed by Gleeson et al. (2012) [59], and the crops highlighted in Figure 4.6c compare well with the groundwater depletion responsibilities found by *Dalin et al.* (2017) [33]. The $WD_{gw}$ found in the North China Plain also compares well with the over-exploitation ratio proposed by *Shi et al.* (2011) [122].

### 4.5.1 Water debt of single crops

We observe important differences in $WD$ across crops and among their top-producing countries (Figure 4.7). On average, global wheat production sustainably relies on soil moisture and surface water, but the groundwater use is unsustainable both in China (2.2 yr) and India (1.5 yr), which together account for over 30% of the global production. Global rice production also contributes to the over-exploitation of aquifers, particularly in China and India. Indian groundwater use (46 $km^3$) is unsustainable with a $WD$ of 1.5 yr, Chinese groundwater use is much more unsustainable ($WD$ of 7.5 yr) even though the groundwater use is half of that in India. In this case, the water use sustainability is related to the different water availability and renewability. Another 10% of rice production is located in Indonesia, where the water use is sustainable thanks to the large fraction of green water use (soil moisture accounts for over 75% of total use). Surprisingly, wheat and rice production in India generate

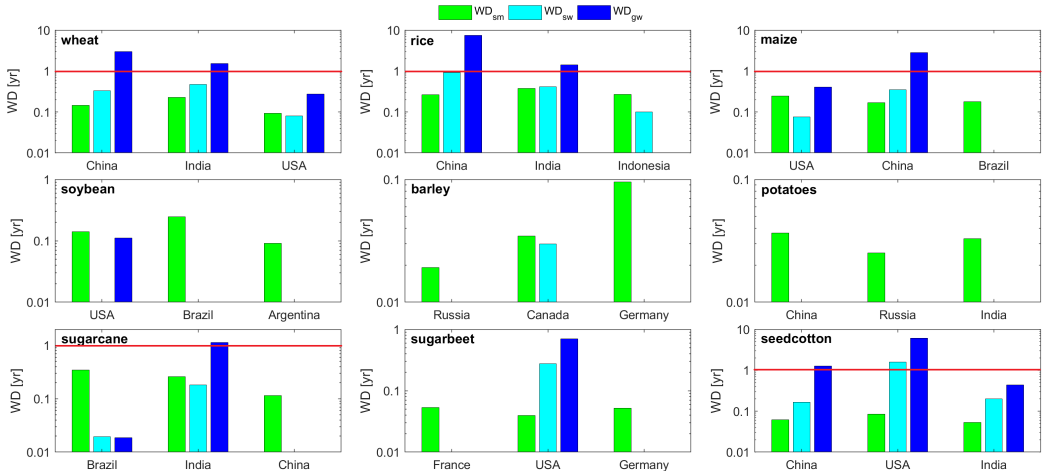


Fig. 4.7 Country-values of the water debt ($WD$) arisen by crop production in the top three producing countries for each of the nine crops. The national level WD values have been obtained as production-weighted averages of the gridded $WD$ values for each water source (equation 4.5). Note the different (logarithmic) scale of the y-axis in each panel.

the same $WD_{gw}$ although rice uses nearly twice as much groundwater as does wheat. This highlights the uneven water availability within the country and the importance of choosing the location of crop production based not only on the soil suitability and climatic conditions, but also on the water availability. Over 60% of global maize production is located in the United States, China, and Brazil. The United States produce twice more maize than China with a three-times larger groundwater use per unit weight, but generate a lower debt with the ground water resources (0.5 yr versus 2.8 yr), highlighting the different renewability rates of aquifers in the two countries. However, despite the sustainability on a national average, maize production is responsible for major $WD$ hotspots over the California Central Valley ($WD$=1.6 yr) and the High Plain aquifer ($WD$=2 yr). Soybean, barley, and potatoes production is completely sustainable in the top-three producing countries: i.e., $WD$ is lower than 0.3 year for all sources, because production is almost entirely rain-fed (Figure 4.3(c)) and sufficient soil moisture is available during the growing season to satisfy the crop water requirement. Notably, Russia, Canada, and Germany produce nearly the same amount of barley, but with very different impacts on the soil moisture. Indeed, Russian barley production generates a $WD$ (number) much lower than those of Canada (number) and Germany (number), despite

a water use per ton of crop (2500 $m^3 \cdot ton^{-1}$) five times larger than those of Canada and Germany. Similarly, sugar crops production is sustainable overall, except for the groundwater use in India that generates a $WD$ of 1.2 yr. The sustainability of sugar crops' water use in Brazil is ensured by the reliance of sugarcane on soil moisture (nearly 90%). Finally, over 40% of cotton production, located in China and the US, is not sustainable and deplete aquifers. Particularly, in the US the $WD_{gw}$ is close to 7 yr due to the large fraction of groundwater use with respect to the total volume, because of insufficient soil moisture availability during the growing season. The water debts arisen in different countries are due to both the domestic and foreign demand for crops.

### 4.5.2 Discussion

The water debt is a physically-based, crop-specific, and locally-defined indicator, assessing the impact of agricultural water use on renewable water resources. By definition, the $WD$ depends on (i) the water use efficiency ($CWF$), (ii) the volume of water used from different sources ($WF_{sm,sw,gw}$), and (iii) the renewability rate of each water source. With respect to traditional assessments of $WF$, the $WD$ enables to discriminate between two countries that, for example, consume the same amount of water to produce a given crop, but with different impacts on the hydrological cycle depending on the type and location of the water source used. For instance, India and the US produced nearly the same amount of wheat (calculated as an annual average over the period 1996-2005), both with an average $CWF$ of 1700 $m^3 \cdot ton^{-1}$, but Indian production generated an average $WD_{gw}$ of 1.5 yr on groundwater resources, which is 5 times larger than that in the US (see Figure 4.7). This happens because thirty percent of Indian production relies on groundwater, while only 5% of American production is sustained by aquifers. However, these two countries show an opposite behaviour when cotton production is considered: i.e., cotton $WD_{gw}$ in the US is close to 8 yr, while it is about 0.5 yr in India, despite American cotton being less water intensive than Indian cotton ($CWF$ of 4300 $m^3 \cdot ton^{-1}$ vs 8500 $m^3 \cdot ton^{-1}$). Such an inversion of impacts is due to both the larger production of cotton in the US and the relatively lower soil moisture availability than in India during the growing season, which increases the irrigation requirement. Water debts quantification gives insights on water

depletion responsibilities and identifies critical points where interventions are more urgent. Long-term changes of the Aral Sea level, due to the diversion of the Amu Darya and Syr Darya rivers, have been shown by *Pekel et al.* [103]. We find that the highest $WD$s along the southern inflow of the Sea are mostly caused by cotton production that over-exploits the local surface water resources, with some cells depleting the resource 40 times faster than it is recharged (Figure 4.8). Crop production in Turkmenistan shows an average $WD_{sw}$ of 21 yr. The Northern inflow of the Aral Sea (i.e., Sir Darya River) is also over-exploited for agricultural production, but lower $WD$s are found because surface water use is lower given that it is mostly provided to rice, which is less water intensive than cotton (i.e., 2560 $m^3 \cdot ton^{-1}$ versus 5455 $m^3 \cdot ton^{-1}$). Both rice production in Kazakhstan and cotton production in Turkmenistan could reduce their $WD_{sw}$ by cutting the crop water use to benchmark values [90]. In fact, cotton production in Turkmenistan uses 40% more water per unit weight than the global average, while rice production in Kazakhstan uses 80% more than the global average. However, cutting the water footprint is not always the most appropriate solution. As shown in the following example of wheat production in China, the local (un)availability of renewable water resources can make a very efficient use of water still unsustainable. Indeed, some provinces where water use efficiency is already high compared to both the country and the global average (Figure 4.9), still show large $WD$s. For example, in the Xinjiang province, crop water-efficiency is very high (i.e., 460 $m^3 \cdot ton^{-1}$), but groundwater use is largely unsustainable ($WD_{gw}$=4.7 yr) due to the low recharge rate of the aquifer owing to scarce precipitation (less than 100 $mm \cdot yr^{-1}$). Similarly, in the North of the Ningxia province (Figure 4.9(b)), water use for wheat production is unsustainable ($WD_{sw}$=1.1 yr) even though the water use efficiency (600 $m^3 \cdot ton^{-1}$) is better than the national average. In the case of Ningxia province, a way to reduce $WD$ could be to transfer the wheat cultivation from North to South East where runoff is much higher, or to increase the production e.g., in the Eastern provinces where soil moisture is widely available during the growing season. Nevertheless, in making these considerations the availability of arable land and adequate labour force should also be taken into account.

## 4.6 Concluding remarks

The water debt repayment time quantifies the mismatch between water use and availability. It provides results detailed by water source and crop, at the grid cell and basin scale; allowing for comparisons among different crop types, different water sources, across countries and water basins. Hence, its analysis may provide useful insights for the resource planning and management in critical areas. Importantly, the $WD$ allows to distinguish two countries showing the same water use efficiency and producing nearly the same amount of crop, but which can incur different water debts with the hydrological cycle, depending on the local availability of the source used for production.

The results obtained with the $WD$ indicator highlights the locations and typology of threats imposed by agricultural production on water resources. Therefore, it can be a tool to monitor food- and water-related targets of the Sustainable Development Goals [13], such as the sustainable food production systems, resilient agricultural practices (Goal 2) and responsible consumption and production patterns (Goal 6). In fact, it may enable informed policy and decision-making for a future agricultural water use that can be locally and globally more sustainable. It is also worth noticing that, in general, water debts are due to the domestic, but also to the foreign, demand for crops, with international food trade playing a role in the increase or decrease of water resources exploitation. The $WD$s will thus be useful in characterizing the sustainability of the trade of water embedded in food trade.
The $WD$ indicator enables to fairly compare the sustainability of crop production related to water resources across regions, based on the local information of water availability and different source renewability. Other water-consuming sectors (e.g., domestic and industrial) should be considered in future studies to complete the picture of sustainability. [20]. Finally, measuring the water use sustainability through the repayment time can help connecting and integrating water resource management with other environmental issues, such as the ecological and the carbon footprint [75]. It is also worth noticing that, in general, water debts are due not only to the domestic, but also to the foreign demand for crops, with international food trade [132, 33] playing a role in the harshening or loosening of water resources exploitation. The water debt will thus be useful in characterizing the sustainability of the trade of water embedded in

food trade. The $WD$ indicator enables to fairly compare the sustainability of crop production related to water resources across regions, based on the local information of water availability and different source renewability. We note that this study only focuses on a portion of agricultural production (which provides about 70% of the global caloric content and 66% of the proteins in global human consumption [34]). Other crops, as well as other water-consuming sectors (e.g., domestic and industrial) should be considered in future studies to complete the picture of sustainability in the assessment of whether current human water use is "meeting the needs of the present generation without compromising the ability of future generations to meet their own needs" [20].

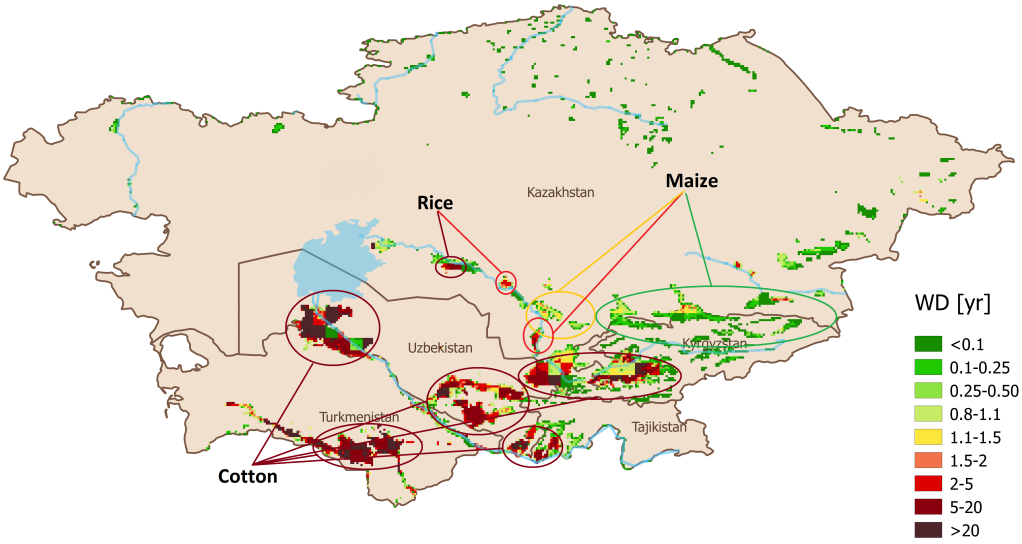


Fig. 4.8 Surface water debt ($WD$) associated with the nine study crops in the Amu Darya and Syr Darya river basins. The color of each circle corresponds to the average $WD$ in the circled area, and the linked crop name indicates the crop mostly responsible for the water debt in each area.

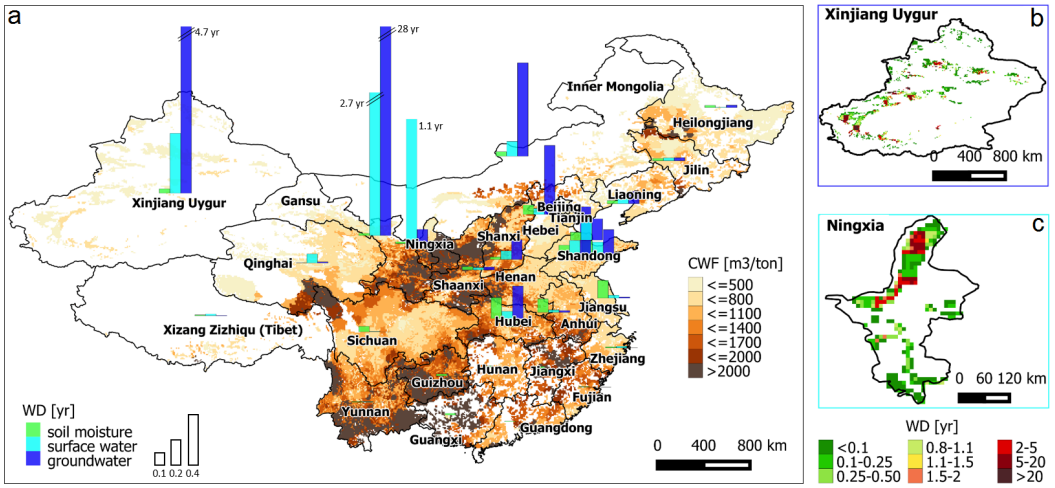


Fig. 4.9 Gridded crop water footprint (volume of water per ton of crop) of wheat in China and histograms of average water debt ($WD$) at the province level (a). Gridded groundwater debts in the Xinjiang Uygur province (b), and gridded surface water debts in the Ningxia province (c).

# Chapter 5

# Link prediction in the virtual water trade network

*The work described in this chapter has been partially derived from paper [137].* International trade of agricultural products has increased substantially in recent decades, due to both food surpluses in some countries [57] and to the central role of import strategies in other countries where food demand overcomes supply [107]. Through the international trade of agricultural goods, water resources that are physically used in the country of production are 'virtually' transferred to the country of consumption. The volume of water virtually transferred is known as 'virtual water' [6, 10] and the transfer is called virtual water trade. The virtual water trade has been studied by a remarkable numbers of studies mostly assessing the water saving and water inter-dependence behind the international exchange of agricultural products. Recently, the analyses of virtual water trade through the tools of complex network theory [99, 15] have also attracted much attention. The idea has been to depict the virtual water trade as a network where countries play the role of nodes and links describe the import/export relations between any two countries (i.e., virtual water trade network, $VWTN$). Important insights about the global architecture of the $VWTN$ have been provided by Konar et al. [81]: their (data-based) analyses quantified the topology of international trade, providing evidence for the existence of the weighted rich club phenomenon and uncovering a global trading hierarchy in which dominant nations connect most peripheral portions of the network. Suweis et al. [130] developed a simple model that captures

the key features of the network by assuming as sole controls each country's gross domestic product and yearly rainfall on agricultural areas, Tamea et al. [132] identified population, gross domestic product, and geographical distance of countries as major drivers of virtual water fluxes. Such studies highlighted the dynamical and intermittent behaviour of the network where a number of links are created and dismissed every year [22, 23]. In spite of the growing efforts devoted to unfold the virtual water network structure and dynamics, an unexplored problem is to understand the single association between any pair of countries involved in the international trade: this can be formalized as a *link prediction task*. Link prediction is the problem of predicting the existence of a link between two nodes, based on the attributes of nodes and/or on the network topology [58]. This Chapter addresses the specific problem of predicting the existence of a link between any two nodes over a directed and weighted network, without any prior knowledge of the network topology, but only knowing the country attributes (e.g., population, gross domestic product, water demand) and the link characteristics (e.g., geographical distance). The proposed methodology allows us to unfold the drivers of link activation and deactivation along the period 1986-2011. The first part of the Chapter describes in details the methodology adopted to predict the link existence between any two countries in the network of virtual water trade. Starting from the assumption of having links between any two countries, we estimate the associated virtual water flows by means of a gravity-law model using country and link characteristics as drivers. We consider the links with estimated flows higher than 1000 $m^3/y$ as active links, while the others as non-active links. Flows traded along estimated active links are then re-estimated using a similar but differently-calibrated gravity-law model. The second part of the Chapter reports an application of the link prediction methodology and provides the accuracy of the methodology both at global and country scale through the evaluation of the associated error. Finally, in the last section, the proposed methodology is put in the context of complex network theory by comparing its performance with other link prediction algorithms available in literature.

The model is able to correctly predict 84% of the existing links observed in year 2011 (16% of the existing links are "missed"), and 93% of the non-existing ones (7% of the non-existing links are spuriously introduced in the network). Although the number of missed links might seem large, the associated

virtual water flow is negligible (1% of the global flow); the links lost by the model, in fact, are mainly those where a minimum volume of virtual water is exchange. We found that geographical distances between countries, population, and fertilizer use are the fundamental drivers of link activation/deactivation. Binary information (e.g., common languages, common religions) and regional trade agreements are also important, while the other variables, pertaining economical and productive aspects, do not give a significant contribution to the link prediction.

## 5.1 The global Virtual Water Trade Network

In the virtual water trade network, each country participating in the food trade is represented by a node. Links between nodes are directed, discriminating the direction of trade flows, and weighted by the volume of virtual water embodied in the traded commodities. The virtual water volume is contributed by a green and a blue component, which refers to the precipitation stored in the soil and vegetation and to the water used for irrigation, respectively. The network is built using information on the trade of agricultural goods between all nations and the virtual water content of each good produced by the nations. The bilateral trade flows are given by the FAOSTAT database (http://faostat.fao.org/), while country-specific virtual water contents for crops and animal products are provided by [89, 91]. Given all virtual water contents and considering exporting countries as producing countries, trade data for each product are converted into virtual water data and summed up over all goods to obtain the total virtual water transfer between trade partners in a given year. Further details about the matrix construction are available from [23] and [132]. The virtual water trade data of each year (from 1986 to 2011) are organized in a matrix, $\mathbf{F}$, whose $F_{i,j}$ element represents the virtual water flux (or link weight) from country $i$ to country $j$. This means that rows represent exporting countries, whereas columns stand for importing countries. The matrix is non-symmetrical because of the network directionality, that is flux from $i$ to $j$ is different than flux from $j$ to $i$. The adjacency matrix, $\mathbf{A} = (a_{i,j})$, associated to the virtual water network is a binary matrix in which the entry $a_{i,j} = 1$ if $i$ and $j$ are directly connected, $a_{i,j} = 0$ otherwise. The number of countries, $n$, and thus

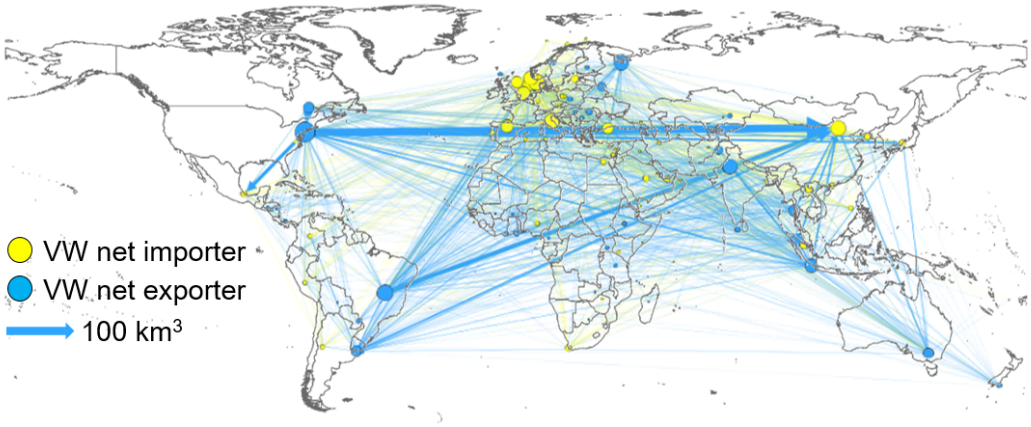


Fig. 5.1 Representation of the virtual water trade network in year 2011. Links are directed from the node of export to the node of import, the link size is proportional to the traded amount of virtual water: e.g., the link from the US to China trades 100 $km^3$ of virtual water. Node colors discriminates net importing (i.e., yellow) and net exporting (i.e., blue) countries while the node size is proportional to the export degree, which expresses the number of exporting relation of each country. Link colors is determined by the color of the sourcing node.

the matrix size, changes from year to year according to political-administrative arrangements (e.g., the collapse of USSR). Figure 5.1 shows the virtual water network in year 2011, associated with the international trade of agricultural goods.

## 5.2 Link prediction methodology

The link prediction algorithm predicts the links between any countries involved in the international virtual water trade, only considering the country and link characteristics (such as population, gross domestic product, geographical distances, etc.), without any prior knowledge of the network structure and topology. Figure 5.2 outlines the overall methodology by providing an illustration of the main steps involved. In STEP 1 (Figure 5.2), we assume that every country is linked with all the other countries with both an import and an export trade relation (i.e., bilateral trade, with $a_{i,j} = a_{j,i} = 1$). Links between nodes belonging to $S_2$ are not taken into account ($a_{i,j} = 0$) due to missing information about node properties; self-connections as well are excluded from our calculations ($a_{i,i} = 0$). We estimate the link weights, $\hat{F}(i,j)$, by means of a

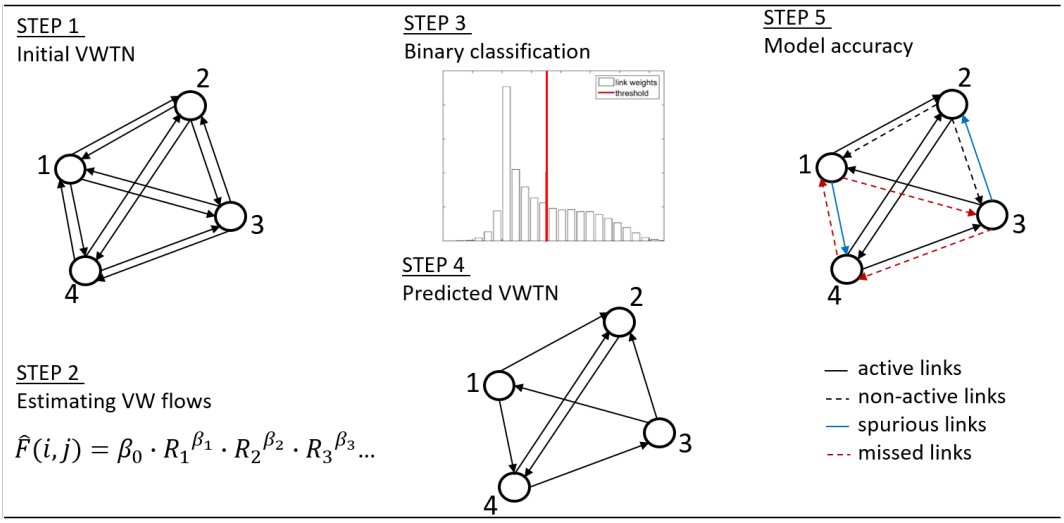


Fig. 5.2 Model approach. Definition of a set of potential links in STEP 1, estimation of the link weights (i.e., virtual water flows) in STEP 2, distinction between active and non-active links through a threshold-based classification in STEP 3, prediction of the $VWTN$ structure in STEP 4, evaluation of the model accuracy in STEP 5.

gravity-law model (STEP 2), using the node-specific and link-specific drivers. Then, in STEP 3, the links with an estimated flow, or weight, higher than a threshold value (highlighted by the red line) are labelled as active links, while the others as non-active links. The predicted active links of the $VWTN$ are shown at STEP 4, while at STEP 5 we asses the model accuracy comparing the real network with the predicted one by differently colouring the links: active links can be correctly predicted (black solid lines) or incorrectly predicted (blue solid lines, “spurious"); non-active links as well can be correctly predicted (black dashed lines) or incorrectly predicted (red dashed lines, “missed").

### 5.2.1 The gravity-law model

The gravity-law model has often been used to study bilateral international-trade flows [102, 56]. Its original formulation, inspired by Newton’s gravity equation, states that the total trade between any two countries is directly proportional to the product of country masses and inversely proportional to their geographic distance [9, 12]. From an empirical perspective, the basic gravity-law has been expanded from its original definition to improve the fit by taking into account

country or trade characteristics that may influence bilateral trade flows in addition to masses and distance [102, 42].

Here, we adopt the gravity-law model to get estimates of the virtual water flows associated to the trade of agricultural goods. Given the complexity of the virtual water network, specific laws describing the virtual water import and export of each country are necessary since a global relationship describing all the exchanged fluxes proved to be inadequate [132]. The import-law, $\hat{F}_{imp}(i,j)$, describes the trade flow from country $i$ to country $j$ as a function of source characteristics ($i$ denotes the exporting country), while the export-law, $\hat{F}_{exp}(i,j)$, expresses the trade flow as a function of destination characteristics ($j$ denotes the importing country), i.e.

$$\hat{F}_{imp}(i,j) = \beta_{0,j} \cdot R_{i,1}^{\beta_{1,j}} \cdot R_{i,2}^{\beta_{2,j}} \cdot R_{i,3}^{\beta_{3,j}} \cdot R_{i,4}^{\beta_{4,j}} ..., \tag{5.1}$$

$$\hat{F}_{exp}(i,j) = \beta_{0,i} \cdot R_{j,1}^{\beta_{1,i}} \cdot R_{j,2}^{\beta_{2,i}} \cdot R_{j,3}^{\beta_{3,i}} \cdot R_{j,4}^{\beta_{4,i}} ..., \tag{5.2}$$

where $\mathbf{R}$ contains the drivers (Section 5.2.2) and $\beta = (\beta_0, \beta_1, \beta_2, ...)$ is the matrix of the model parameters.

We fit bilateral-trade flows along all the assumed links (STEP 1, Figure 5.2) managing the gravity-law equations as linear multivariate regressions between the logarithm of fluxes and the logarithm of drivers. In order to avoid problems with the logarithm of null fluxes (which are those corresponding to non-existing links), we assign to them a fictitious virtual water flux of 1 $m^3$ per year; this assignment does not compromise the original matrix, $\mathbf{F}$, where the smallest flux is around 70 $m^3$ in year 2011 and the average flux is about $1{\cdot}10^7$ $m^3$/year. Model parameters are interpreted as regression coefficients and estimated with the ordinary least square method. Significant variables are then identified applying a Student's $t$-test with a 5% significance level. Finally, new regression coefficients are determined using only statistically-significant variables; when none of the variables is significant, only the $\beta_0$ coefficient is taken as the modelled flux.

When the flow from $i$ to $j$ is between nodes that belong to subset $S_1$, the flow is fitted by both the export law of $i$ (equation (5.2)) and the import law of $j$ (equation (5.1)) because both nodes have all the drivers available, and a mean flow, $\hat{F}(i,j)$, is then calculated by averaging the import, $\hat{F}_{imp}(i,j)$, and the export, $\hat{F}_{exp}(i,j)$, estimates. Averaging the flows obtained with the two laws

Table 5.1 Possible drivers of the gravity-law model, with their source and spatio-temporal availability.

| | Label | Description | Source | n° of countries | Time interval |
|---|---|---|---|---|---|
| 1 | $P$ | Population | FAOSTAT (http://faostat.fao.org/) | 228 | 1961-2011 |
| 2 | $wd$ | Virtual water of national consumption in $m^3$/cap | [22, 23, 132] | 175 | 1986-2011 |
| 3 | $Pr$ | Annual on cultivated area | Our computation on GAEZ dataset and [98] | 182 | 1960-2000 |
| 4 | $ET_0$ | Annual evapotranspiration on cultivated area | Our computation on GAEZ dataset and [98] | 182 | 1960-2000 |
| 5 | $A$ | Arable area in ha/cap | FAOSTAT | 219 | 1960-2000 |
| 6 | $A^I$ | Area equipped for irrigation in ha/cap | FAOSTAT | 184 | 1960-2000 |
| 7 | $N$ | Nitrogen fertilizers in tonnes of nutrients | FAOSTAT | 160 | 1986-2012 |
| 8 | $K$ | Potash fertilizers in tonnes of nutrients | FAOSTAT | 155 | 1986-2012 |
| 9 | $Ph$ | Phosphate fertilizers in tonnes of nutrients | FAOSTAT | 154 | 1986-2012 |
| 10 | $AP$ | Agricultural population | FAOSTAT | 154 | 1986-2012 |
| 11 | $GDP$ | Gross domestic product in $/cap | http:// unstats.un.org/unsd/snaama/dnlList.asp | 208 | 1970-2011 |
| 12 | $AV$ | Agricultural value of production | FAOSTAT | 205 | 1986-2013 |
| 13 | $wp$ | Virtual water of agricultural production in $m^3$/cap | [22, 23, 132] | 216 | 1986-2011 |
| 14 | $vwv$ | Virtual water value in $/$m^3$ | This study | 150 | 1986-2011 |
| 15 | $rta$ | Regional trade agreements | CEPII (http://www.cepii.fr) | 208 | - |
| 16 | $D$ | Distance | CEPII | 255 | 1986-2011 |
| 17 | $b$ | language, colony relation, contiguity, religion | Our computation on CEPII dataset | - | - |

results in an improvement over the single law. Otherwise, if the flow is from node $i$, belonging to $S_1$, to node $j$, belonging to $S_2$, (or *vice versa*) the flow is fitted only with the import or export law of node $j$ by taking into account all the properties of node $i$. In this case, the export law of node $i$ cannot be estimated because $j$ lacks of the drivers to set equation (5.2). Finally, if both $i$ and $j$ belong to $S_2$, the flow is not estimated by the model. In this latter case, since we are not able to estimate the link weight, we cannot predict the existence of the link. However, it is worth noting that the flow traded along links connecting nodes in $S_2$ only constitutes 0.5% of the global virtual water flow (see Table 5.2), thus allowing to model almost the total volume of virtual water trade. Moreover, modelling links between $S_1$ and $S_2$ allows to cover the 96.2% of the flows to/from countries in $S_2$.

### 5.2.2 Possible drivers of the virtual water trade

We analyse as possible drivers of the virtual water trade 17 factors (or drivers) variably determining food demand/supply, water availability, and trade flows. We consider both node-specific and link-specific country attributes.

Node-specific attributes affect food demand, water availability, agricultural productivity potential, and economic power of each country. In particular, *Food demand* is determined by population ($P$) and per-capita demand of water embedded in agricultural goods ($wd$). *Water availability* depends on climatic conditions, that are described here by annual rainfall ($Pr$) and cumulative reference evapotranspiration ($ET_0$) over cultivated areas. *Productivity potential*

is influenced by per-capita arable area ($A$), per-capita irrigated area ($A^I$), use of fertilizers (such as nitrogen ($N$), potassium ($K$), and phosphate ($Ph$), and agricultural population ($AP$); *Economic power* depends on per-capita gross domestic product ($GDP$), agricultural value ($AV$) which measures the output of the agricultural sector, virtual water value ($vwv$), virtual water of per-capita agricultural production ($wp$), and trade relations such as the regional trade agreements ($rta$).

Trade relations are also strongly impacted by link-specific attributes, such as the geographical distance ($D$) between two nodes and aggregated binary information ($b$), including common official language, colony relation, contiguity, and common religion.

Sources and spatio-temporal availability of the variables adopted in this study are given in Table 5.1, while details on data characteristics and pre-processing are provided as on-line Supplementary material. The spatial coverage of data is variable over the studied period, as every year there are few countries lacking one or more drivers. Therefore, we group the countries in two subsets: the first one ($S_1$) contains countries that have all the variables available in the studied years, while the other one ($S_2$) contains the countries that lack one or more variables. Properties of the $n_1$ nodes from subset $S_1$ are reported in a $n_1$ x 17 matrix, $\mathbf{R}$, while all properties of subset $S_2$ are left out from the analyses because they are incomplete. By doing so, we ensure that all the link weights are estimated with exactly the same data sources.

**Food demand.** Food demand is determined by population ($P$) and per-capita demand of water embedded in consumed agricultural goods ($wd$). The population for each country and for each year is taken from the FAOSTAT database (http://faostat.fao.org/). The mean values of the per-capita water demand for each country in the period 1996-2005 are taken from Mekonnen et al. [92]. The water demand represents the water footprint of national consumption and it is defined as the total volume of water that is used to produce the goods and services consumed by the inhabitants of the nation. It consists of two components: the internal and external water footprint of national consumption [92]. The internal water footprint expresses the use of domestic water resources to produce goods and services consumed by the country's population; the external water footprint is defined as the volume of water resources used in other nations to produce goods and services consumed by the population in

the nation under consideration. The water demands accounts for the green, blue, and grey water consumption.

**Water availability.** Water availability depends on annual rainfall ($Pr$) and cumulative reference evapotranspiration ($ET_0$) over cultivated areas. The values of $Pr$ and $ET_0$, expressed in mm/month, are available from CRU CL 2.0 [98] at a global scale at 10 arc minute latitude/longitude resolution. We aggregated these values at the country scale by a weighted average of the annual values, using as weights the cultivated areas taken from the GAEZ database (http://www.gaez.iiasa.ac.at/).

**Productivity potential.** Productivity potential is influenced by the per-capita arable land ($A$) and the per-capita irrigated land ($A_i$), the use of nitrogen ($N$), potassium ($K$), and phosphate ($Ph$), and the agricultural population ($AP$).

The arable and irrigated lands are available for each country and each year on the FAOSTAT database. The arable area represents the area cultivated with temporary agricultural crops, temporary meadows for mowing or pasture, and land temporary fallow (less than five years). The irrigated land is the area equipped for irrigation that can be totally or partially irrigated in a specific year. Both $A$ and $A^I$, expressed in hectares, are divided by the population to obtain the per-capita values.

Fertilizers are available from the FAOSTAT database as well. In particular, we focused on the consumption of nitrogen, potassium, and phosphate expressed as tonnes of nutrients consumed in each country for each year along the study period; they represent proxies for the agricultural efficiency and high crop yields. The agricultural population represents the population employed in agriculture in each country across the study period.

**Economic power.** Economic power is related to the gross domestic product ($GDP$), the agricultural value of production ($AV$), the per-capita water production ($wp$), the virtual water value ($vwv$), and the regional trade agreements ($rta$).

Per-capita gross domestic product for each country and year is available from the National Accounts Main Aggregates Database (http://unstats.un.org/unsd/snaama/dnlList.asp). Further details are available in [132]. Per-capita agricultural value of production represents the gross production value expressed in US$ divided by the population for each year and country; it is a measure of the production in

monetary terms at the farm gate level.
The virtual water of agricultural production of each country and year is evaluated by multiplying the quantity (expressed in tons) of each agricultural good produced in a country by the national virtual water content (expressed in $m^3$/ton), and summing up over all good to have the national virtual water volume of the agricultural production. For more details, refer to Tamea et al. [131].
The virtual water value stands for the price of the virtual water used for the production of agricultural goods; it is the ratio between the gross domestic cost of the agricultural production (expressed in US$), provided by FAOSTAT, and the national the virtual water of per-capita agricultural production, expressed in $m^3$, provided by Tamea et al. [132].
The number of regional trade agreements for each country is available from the World Trade Organization database (http://rtais.wto.org/UI/PublicMaintainRTAHome.aspx).
**Link-specific attributes** are those characterizing a link instead of a node and include the geographical distance between countries and the binary information. The distance between countries, which is taken from the CEPII dataset (http://cepii.fr/CEPII), is organized in a matrix of geographical distances measured between the most populated cities of the countries.
The binary information, $b$, can be see as a multiple-dummy variable, derived from the aggregation of four different types of binary information, including common official language, colony relation, contiguity, and common religion, derived from the CEPII database. First, we construct a time-invariant (1-10) matrix for each dummy variable, $b_x$ with $x = 1, 2, 3, 4$. We set the matrix values to 1 instead of 0 and to 10 instead of 1 in order to avoid problems with the logarithms. Then, we aggregated the dummy variables in a single matrix, $b$, according to equation (5.3). The elements of $b$ run from 1 (no relation between the nodes) to 10000 (four different relations between the nodes).

$$b = b_1 \cdot b_2 \cdot b_3 \cdot b_4 \tag{5.3}$$

### 5.2.3 Threshold evaluation

The proposed methodology grounds the link activation between any two countries ($i$ and $j$) on the link weights estimated by the gravity-law model (equa-

tions (5.1,5.2)). The frequency distributions of the weight estimations expressed in logarithm are shown in Figure 5.3(a), where the light grey histogram refers to the non-existing links and the dark grey histogram refers to the existing links. In order to efficiently distinguish among active and non-active links, a binary classifier (or threshold) is introduced: links are classified as active if their estimated weight is higher than the threshold, non-active otherwise. Depending on the threshold values (the blue, red, and yellow lines in panel (a) indicate some examples of threshold values), the model correctly identifies a number of existing or non-existing links. In order to determine the threshold with the best accuracy both for existing and non-existing links, we evaluate the true positive rate or sensitivity, $TPR$, and the false positive rate or specificity, $FPR$, for every possible threshold value in the flux domain, according to

$$TPR = \frac{TP}{TP + FN}, \tag{5.4}$$

$$FPR = \frac{FP}{TN + FP}. \tag{5.5}$$

The true positive rate (equation (5.4)) is determined by the number of true positive ($TP$) that refers to the existing links correctly Identified by the model and the number of false negative ($FN$) that refers to the links which are “missed" by the model because their estimated weights fall under the threshold. The false positive rate (equation (5.5)) depends on the number of false positive ($FP$), namely the “spurious" links that the are erroneously classified as active links even though they are non-existing, and the number of true negative ($TN$) that indicates the correctly identified non-active links. The sum of the $FPR$ (that can be seen as the percentage of “spurious" links) and the false negative rate, i.e., $FNR = 1 - TPR$, (which is the percentage of “missed" links) quantifies the link prediction error, $E$, associated to the proposed methodology, that reads

$$E = \frac{FP}{TN + FP} + \frac{FN}{TP + FN} \tag{5.6}$$

The graphical plot of the $TPR$ and $FPR$ values for the three threshold values (namely,1.8, 3.0, 4.0) shown in panel (a) with the blue, red, and yellow lines, are marked in panel (b), where the resultant dashed curve is called ROC

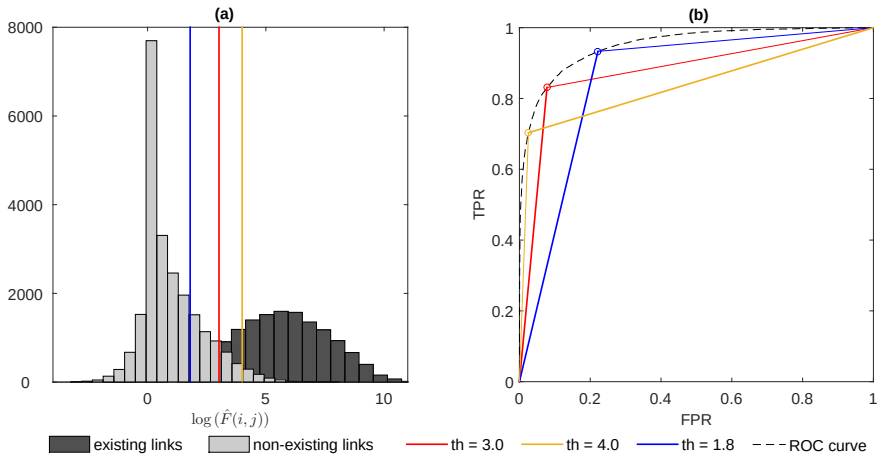


Fig. 5.3 Threshold evaluation. (a) Frequency histograms of the estimated weights, expressed in logarithm, associated to existing (in dark grey) and non-existing (in light grey), and three examples of threshold logarithmic values (that are 1.8, 3.0, 4.0) in blue, red, and yellow. (b) The dashed Receiving Operating Characteristic curve illustrates the performance of the binary classifier as the threshold is varied; each point along the curve indicates the True Positive Rate, $TPR$, and the False Positive Rate, $FPR$, of a threshold that can be selected in the flux domain. The blue, red, and yellow points represent the classifier accuracy for the three thresholds shown in panel (a).

curve (i.e., Receiver Operating Characteristic curve). The area under the ROC curve is a measure of the correct classification probability; therefore, by comparing the trapezoidal area, $A$, defined by (0,0), ($FPR$,$TPR$), (1,1), (1,0), as shown in Figure 5.3(b) for every possible threshold, we can determine the most suitable value. It must be noticed that the area complementary to $A$ is equal to half of the link prediction error and, thus, maximizing $A$ means minimizing $E$.

The evaluation of the most suitable threshold is repeated for each year from 1986 to 2011 and we find logarithmic values in the range of 2.74 and 3.12, thus we fixed the model logarithmic threshold equal to 3, namely 1000 $m^3$/year.

### 5.2.4 Virtual water flows assessment

Once the structure of the unweighted virtual water network has been reconstructed (see Figure 5.2, STEP 4), we apply the gravity-law equations (5.1,5.2) to assess the virtual water volume traded along (predicted) active-links. We

adopt the same procedure as in STEP 2 (Section 5.2.1), with a modified calibration of the regression coefficients; in fact, now we only fit the trade flows of the predicted active links (or the true positive). Depending on the number of trade partners (i.e., node degree, $K$), the gravity-law equation of each country is calibrated on a different number of flows, as in the network there are countries with very few links (i.e., the most peripheral ones) and other with more than a hundred links (i.e., highly connected nations). Consequently, it may not be possible to fit a complete regression for every country, but, depending on the node degree one can use a most appropriate regression. Selecting a null model and a simplified model (with only 3 regressors) besides the complete model (with 17 regressors), the following function is built:

$$\hat{F}(i,j) = \begin{cases} \beta_0 & K < 10 \\ \beta_0 \cdot R_1^{\beta_1} \cdot R_2^{\beta_2} \cdot R_3^{\beta_3} & 10 \leq K < 25 \\ \beta_0 \cdot R_1^{\beta_1} \cdot R_2^{\beta_2} \cdot R_3^{\beta_3} \cdot ... \cdot R_{17}^{\beta_{17}} & K \geq 25 \end{cases}, \tag{5.7}$$

where $K$ equal to 10 links and 25 links are taken as minimum sets, in the simplified and complete model, to avoid overfitting issues. The thresholds on $K$ have been chosen based on a trial-and-error procedure to find the best balance between accuracy and robustness.
When the flow can be fitted using both gravity-laws (i.e., $i$ and $j$ belong to $S_1$), a mean flux, $\hat{F}(i,j)$, is calculated by a weighted mean of the logarithms of the import ($\hat{F}_{imp}(i,j)$) and the export ($\hat{F}_{exp}(i,j)$) estimated flow, according to

$$\mathrm{Log}(\hat{F}(i,j)) = \frac{\mathrm{Log}(\hat{F}_{imp}(i,j)) \cdot K_{imp}(j) + \mathrm{Log}(\hat{F}_{exp}(i,j)) \cdot K_{exp}(i)}{K_{imp}(j) + K_{exp}(i)}, \tag{5.7}$$

where, $K_{imp}$ and $K_{exp}$ are the vectors of import and export node degrees, respectively. In this way, the flow estimated with the most robust law contributes to the weighted mean to a greater extent.

## 5.3 Application and Results

This section presents the application of the link prediction methodology to the network of virtual water trade. We propose an initial focus on year 2011 with the aim of showing the model accuracy and identifying the major factors driving the link activation/deactivation, followed by the application of the model to the whole period (1986-2011) to show the temporal dynamics of the network drivers. Then, we asses the model accuracy at the country scale to identify the regions where the model works best. Finally, we discuss the application of the gravity-law model to the import and export flows traded along predicted active links to investigate the entities of the exchanged flows along the study period and their main controlling factors.

### 5.3.1 Drivers of link activation and deactivation

#### VWTN in year 2011

In year 2011, the 213 nodes of the $VWTN$ are linked by 16254 links, which exchange 2720 $km^3$ of virtual water, while the number of non-existing links (that could have been active in the previous years or could activate in the future) is 28902.
In order to select the drivers of the link activation/deactivation among all the possible variables (presented in Section 5.2.2) to predict the network configuration, we perform a stepwise regression which involves (i) starting with no variables in the model, (ii) testing the addition of each variable, using as comparison criterion the link prediction error calculated according to equation (5.6), (iii) adding the variable that improves the prediction accuracy the most, and repeating this process until all variables are included.
In this analysis, the subset $S_1$ includes 117 countries having all the 17 drivers available and representing a major portion of the global population (92%) and of the traded virtual water flow (87%), as shown in Figure 5.4(c). The characteristics of the modelled network and the partition of the nodes in subsets $S_1$ and $S_2$ are reported in Table 5.2.

Figure 5.4(a) shows separately the components of the link prediction error for year 2011: the red lines stand for the percentage of “missed" links at each

Table 5.2 $VWTN$ characteristics in year 2011. The partition of the nodes in the subsets $S_1$ and $S_2$ is determined by the drivers availability for each country; $S_1$ contains all the countries having the 17 considered drivers available, while $S_2$ constitutes the complementary subset to $S_1$. The percentage of traded virtual water volume reported in the table refers to the global flow of 2720 $km^3/y$.

| | $S_1$ | $S_1$-$S_2$ | $S_2$ |
|---|---|---|---|
| nodes | 117 | | 96 |
| active links | 8767 | 6688 | 799 |
| non-active links | 4805 | 15776 | 8321 |
| traded volume [%] | 87 | 12.5 | 0.5 |

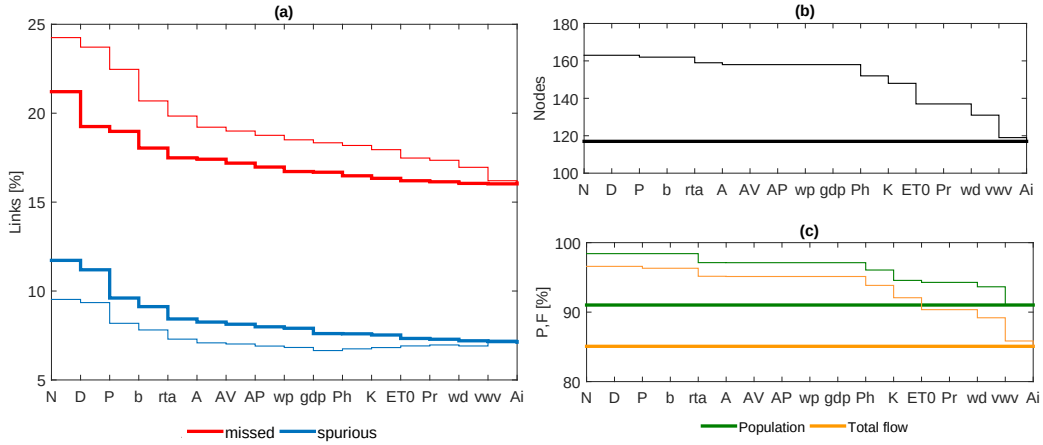


Fig. 5.4 Results of the link prediction exercise for the virtual water trade network in year 2011. (a) The red lines represent the percentage of missed links, while the blue lines stand for the percentage of spurious links; solid lines refer to the case in which the size of $S_1$ is fixed to 117 nodes (those with all the 17 drivers available) and the multivariate regression has always 17 variables, while thin lines are for the case of a variable size of $S_1$, depending on the number of variables considered in the multivariate regression. The abscissa reports the driver sorted as a function of the minimum attainable link prediction error. (b) Number of nodes having all the 17 drivers available (thick line) and number of nodes for an increasing number of drivers (thin line). (c) Percentage of global population and virtual water flow representative of the selected nodes. In all panels, thick lines refer to the case of a complete regression (with all the 17 regressors), whereas thin lines stand for regressions with a variable number of regressors, from 1 to 17.

selection step, while the blue lines are for the percentage of “spurious" links; the abscissa reports the drivers ordered according to their progressive insertion in the regression. It is clear from panel (a) that the progressive insertion of variables improves the model ability to predict links by correctly discriminating between active and non-active links: the percentage of missed links progressively decreases from 22% to 16% (red solid line) and the spurious links decreases from 12% to 7% (blue solid line). Both for missed and for spurious links, the main reduction of the associated error is mostly due to very few variables, such as the use of fertilizers ($N$), the distance between countries ($D$), the population ($P$), the binary information related to the countries ($b$), and the regional trade agreements ($rta$). Distances and binary information allow the most relevant reduction of missed links (as shown by the higher steps of the line), while population is the most relevant driver for spurious links. The other variables, instead, are less incisive in determining a further improvement of model performance. For example, among the 17 possible drivers, 5 variables are sufficient to correctly identify 84% of the existing links and 93% of the non-existing links.

In order to analyse the model accuracy along the study period, we include in the regressions all the 17 drivers to estimate the fluxes and distinguish among active and non-active links. We observe that the percentage of missed links decreases from 19.5% (in year 1986) to 16% (in year 2011), while the percentage of spurious links increases from 4.5% to 7%. The time evolution of the link prediction error, split in its component, is shown in the Supplementary material (Figure S2).

Despite being similar in their behaviour, missed and spurious links are quite different in absolute terms: in fact, considering a 17-variables regression (whose associated error correspond to the last position in abscissa of Figure 5.4(a)), the method loses 2525 active links and adds to the network 1507 spurious links. Although the missed links are more than spurious links, the total virtual water flow traded along them is negligible (1% of the global virtual water flow). The links lost by the model, in fact, are mainly those between small importer or exporter countries, and, thus, the error is considered acceptable.

Once all the drivers have been classified and ordered according to the abscissa in Figure 5.4(a), it is possible to chose a proper subset of variables depending on the portion of the network that one decides to model. For example, if only

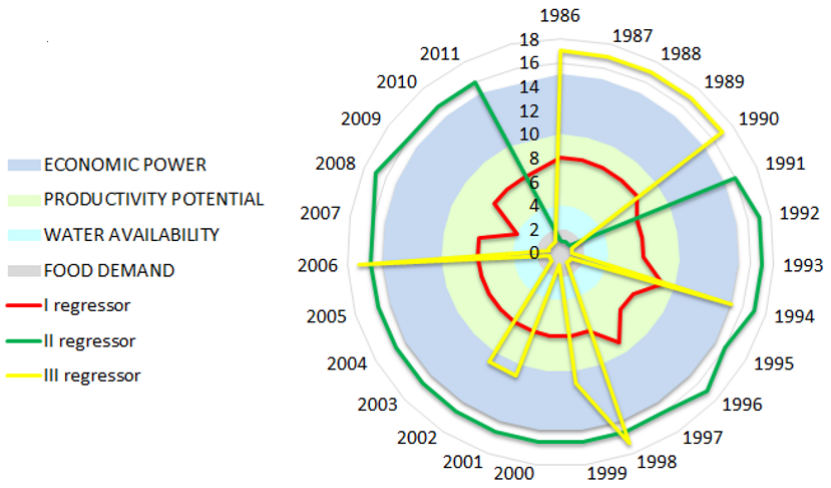


Fig. 5.5 Time evolution of the fundamental variables driving the link activations. Years run in clockwise direction. Node-specific variables are grouped in four classes, starting from the center with *Food demand* (1-2), *Water availability* (3-4), *Productivity potential* (5-10), and *Economic power* (11-15); link-specific variables are those out of the coloured zones, with position 16 and 17. The red line indicates which is the variable that most minimizes the link prediction error when a univariate regression is considered; the green and yellow lines refer to multivariate regressions, with two and three variables, respectively.

the driver $N$ is taken into account, the predicted network will loose 24% of active links and spuriously add 9% of non-active links (according to the red and blue thin lines in Figure 5.4(a)); $N$ is available for 163 countries (Figure 5.4(b)), covering 98% of the global population and 96% of the global virtual water flow (Figure 5.4(c)). Then, by adding other drivers, the model accuracy improves, but the number of considered countries decreases (thin lines in panel (b,c)) according to the definition of $S_1$ given in Section 5.2.2.

### Temporal variability

Considering the temporal variability of the $VWTN$ over the studied period, the stepwise regression described in Section 5.3.1 produces, every year, a different sorting of the drivers. This is due to the different structure of the network, but also to the different spatial coverage of the country-specific variables (see Section 5.2.2). For the sake of simplicity, we report in Figure 5.5 only the sorting of the best three variables.

From 1986 to 2011, years run along the circle in clockwise direction and the drivers, (from 1 to 17, see Table 5.1) are indicated at different radial distance, from the center to the periphery of the circle. Nodes-specific variables are shown grouped in their categories (Section 5.2.2): in position 1 to 2 there are *Food demand* variables ($P$ and $wd$), in 3 to 4 there are *Water availability* drivers ($Pr$ and $ET_0$), *Productivity potential* drivers ($A$, $A^I$, $N$, $K$, $Ph$, and $AP$) are in positions from 5 to 10, and *Economic power* variables ($gdp$, $AV$, $wp$, $vwv$, $rta$) occupy positions from 11 to 15. Finally, link-specific characteristics ($D$, $b$), are in positions 16 and 17, out of the coloured zones. Considering a univariate regression (Figure 5.5, red line), one observes that the most significant variable for link prediction is always the one from the *Productivity potential* category, with the only exception of year 2008, when water availability (i.e., $ET_0$) is more informative. Looking at a bivariate regression (green line), link-specific variables add the most significant information to the univariate regression for most of the years (1991-2011), but the *Food demand* ($P$) is the most informative from 1986 to 1990. Differently from the first and the second regressor, it is more difficult to identify a clear trend over time for the third driver: only from 1986 to 1990 one can note a continuity in the role of link properties, that clearly compensate a lack in the bivariate regression. From this temporal analysis, the economic power of nodes seems not to consistently influence the link prediction accuracy, with only few exceptions.

### Accuracy at the country scale

After having analysed the accuracy of the model at the global scale, results are now presented in terms of percentages of missed and spurious links at the country scale. Figures 5.6(a,b) refer to exporting relations, while Figures 5.6(c,d) refer to importing relations; missed links are in the scale of red, spurious links are in the scale of green. Existing links are well-captured by the model, and countries with a number of missed links in the range of 5-15% prevail in the map of exporting and importing relations. The model predicts very well the links of the Americas, North Europe, Russia and Kazakhstan, South-East Asia, Indonesia and Australia, with a number of missed links generally lower for the exporting relations. Instead, links pertaining to African countries are not well captured (apart from the export of Botswana), especially for exporting relations,

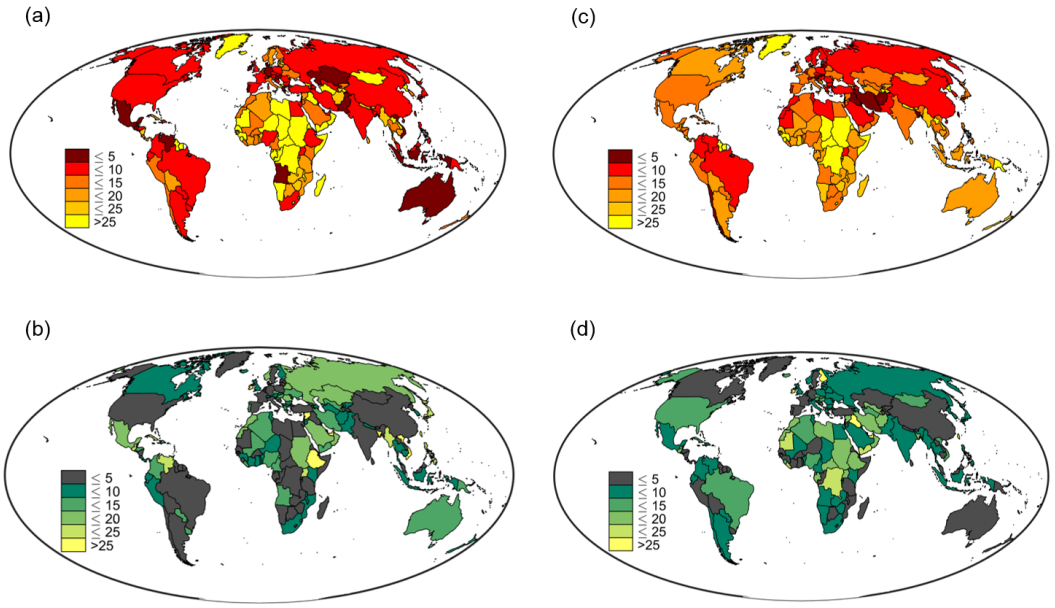


Fig. 5.6 Percentage of missed and spurious links at the country scale, relative to exporting (left) and importing (right) countries. Missed links are in the scale of red, spurious links are in the scale of green.

and results for this region are highly heterogeneous. The scarce performance of the model for African countries may be due (i) to the small number of active links and (ii) to the small volume of the virtual water flows. In fact, with a low number of active relations, the gravity-law coefficients are mostly calibrated on the unitary flows traded along null links, and only partially calibrated on the small virtual flows traded along active links; therefore, the model tends to underestimate the link weights of African countries and a major portion of them falls under the threshold of 1000 $m^3$. This is also a possible explanation to the fact that the model performs better for the African imports than for the exports.

Looking at the percentage of spurious links (Figure 5.6 (b,d)), the model performs better in the identification of non-active links; in fact, the threshold spuriously adds to the real network from 5 to 10% of the null links. It must be noted that some regions where spurious links are lower than 5% (such as some exporting countries in Africa) are those with a corresponding percentage of missed links higher than 25%; the opposite behaviour of the error associated to missed and spurious links reflects the choice of the threshold at the global scale which is suboptimal in minimizing the link prediction error in these regions.

The definition of country-specific (or region-specific) thresholds would enable better predictions at the price of a lower generality and increased need for calibration.

### 5.3.2 Virtual water fluxes: stepwise regression

Here, we adopt the gravity-law model (Section 5.2.1) to fit the virtual water flows traded along the predicted active links and investigate the fundamental drivers of the virtual water fluxes. Similarly to Section 5.3.1, we order the drivers with the help of a stepwise regression using (as a comparison criterion) the adjusted coefficient of determination, $R^2_{adj}$, evaluated at the global scale. At each step, we add to the regression the variable that mostly improves the value of $R^2_{adj}$, namely the model fitting ability. We find that population, distance, and gross domestic product are the most important variables: their progressive insertion in the regression, in fact, improves the value of $R^2_{adj}$ from 0.39 to 0.52 as shown in Figure 5.7. This finding is confirmed by previous studies on the drivers of the virtual water trade [132]. Adding further variables increases the value of $R^2_{adj}$ up to 0.65 and up to 0.90 when we measure an $R^2_{adj}$ weighted with the (non-logarithm) virtual water volumes traded along active links. The remaining virtual water flows, associated to the missed links, constitute only the 1% of the global flow (refer to Table 5.2).

In the inset of Figure 5.7, we show the temporal variability of $R^2_{adj}$ values over the study period, considering two different subsets of drivers: one subset contains all the 17 variables, the other contains only the 5 variables that mostly contribute to improve the model fitting ability (i.e., population, distance between countries, gross domestic product, water production, and water demand). In the case with 17 variables, the value of $R^2_{adj}$ along the studied period reaches 0.65 in year 2011; in the case with 5 variables, the performance is worse with a final $R^2_{adj}$ equal to 0.60. Even if the 17-variables model outperforms the 5-variables one, it is worth noting that the second one is more robust in terms of model parameters, since they are calibrated on a larger number of fluxes and also on a larger number of countries given that $S_1$ is larger in the case of a 5-variables model.

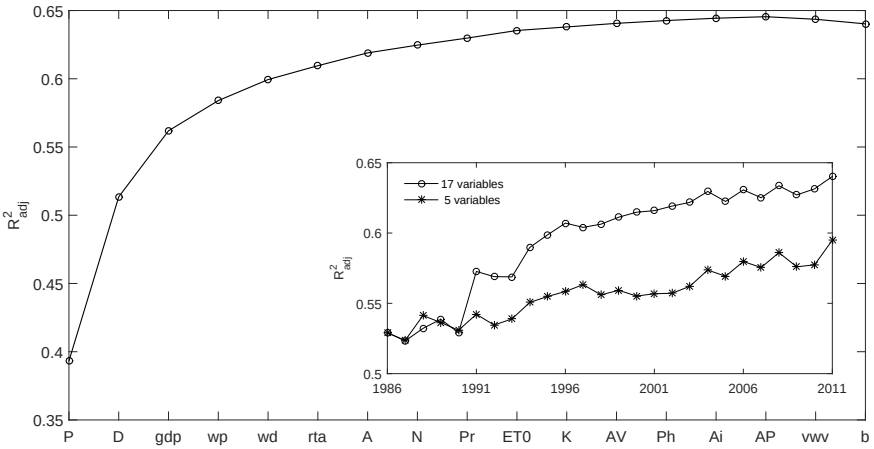


Fig. 5.7 Stepwise selection of the major drivers of the virtual water flows in year 2011 for a network of 117 nodes, by adopting the $R^2_{adj}$ value as the comparison criterion, using a 17-variables regression. The inset reports the temporal trend of the worldwide adjusted coefficient of determination, considering a 17-and a 5-variables regressions.

## 5.4 Discussion

Prior studies have analysed the global architecture of the $VWTN$, highlighting the dynamical and intermittent behaviour of the network [22, 23]: new trade relations keep developing between countries over time. In order to understand the evolution of the whole network, it is necessary to analyse the association between each pair of nodes across time. In the present study, we have addressed this problem as a *link prediction task* in the virtual water network.
The accuracy of the proposed methodology is illustrated by the red ROC curve shown in Figure 5.8, whose points represent the accuracy of all the possible thresholds that can be employed to distinguish among active and non-active links.

In the context of link prediction, it is of interest to compare our methodology with other link prediction algorithms generally used in the field of complex network theory. In particular, we report the ROC curves of four local similarity indices (namely, the Common Neighbors, $CN$, the Salton index, $Salton$, the Hub Depressed index, $HDI$, and the Preferential Attachment, $PA$) that are solely based on the network topology ([87]). These indices (or scores) indicate the similarity between node $i$ and node $j$ (refer to Table 5.3). For example,

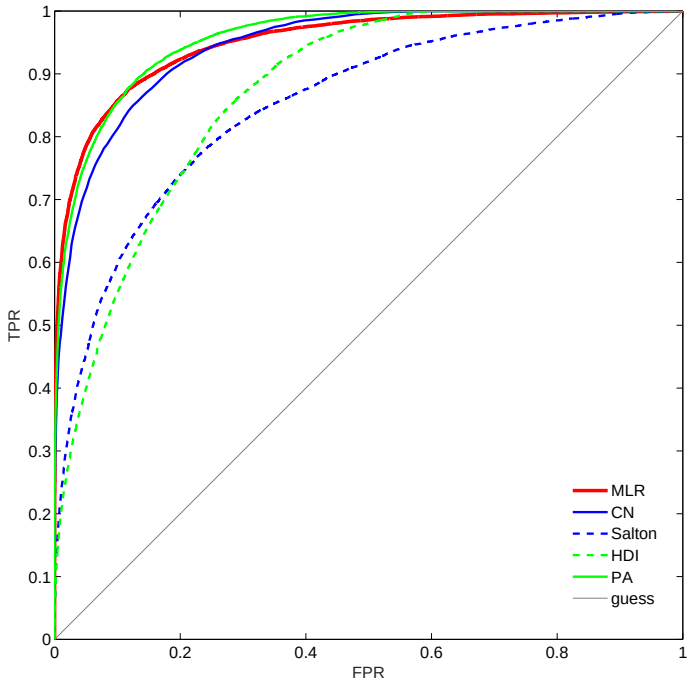


Fig. 5.8 ROC curves to quantify the accuracy of the proposed methodology (i.e., multi-linear regression model, $MLR$, in red) in comparison with other prediction algorithms (i.e., local similarity indices) from the field of complex network theory. The blue solid line stands for the Common Neighbor index ($CN$), the blue dashed line is for the Salton Index ($Salton$), the green solid line indicates the Preferential Attachment index ($PA$), and the green dashed line is for the Hub Depressed Index ($HDI$). Each point along the curves refers to a possible value of the threshold to distinguish among active and non-active links, and it denotes the True Positive Rate ($TPR$) and the False Positive Rate ($FPR$).

Table 5.3 Definition of the link score, as the local similarity between node $i$ and node $j$, according to different indices: Common Neighbors ($CN$), Salton ($Salton$), Hub Depressed index ($HDI$), Preferential Attachment ($PA$).

| | score |
|---|---|
| $CN$ | $\lvert\Gamma(i) \cap \Gamma(j)\rvert$ |
| $Salton$ | $\frac{\lvert\Gamma(i)\cap\Gamma(j)\rvert}{\sqrt{k_i \cdot k_j}}$ |
| $HDI$ | $\frac{\lvert\Gamma(i)\cap\Gamma(j)\rvert}{\max(k_i,k_j)}$ |
| $PA$ | $k_i \cdot k_j$ |

$CN$ indicates the number of trade partners that node $i$ and node $j$ have in common; it is calculated as the intersection between the neighbors (or trade partners) of node $i$ indicated by the set $\Gamma(i)$ and the neighbors of node $j$ ($\Gamma(j)$). All links are then ranked according to their scores, and the links connecting more similar nodes are supposed to be of higher existence likelihood. Indeed, all the links with a value of higher than a threshold are considered active links, while the others are labelled as non-active links.

By comparing the accuracy of our methodology with the one achieved by the local similarity indices, only the Common Neighbours and the Preferential Attachment (whose accuracies are represented by the blue and the green solid curves in Figure 5.8) perform as well as our model. But, despite performing similarly, these indices are based on the network structure that is supposed to be known and, thus, every time we evaluate the score of a link between any two nodes, we need to know all the other relations between the considered nodes and the rest of the network. Differently, our methodology has the great advantage to reconstruct the network architecture without any prior knowledges of the network, since it only requires country attributes. More generally, our study provides a new link prediction framework which can be used for small networks, where the number of nodes is limited and the node and link properties can be easily reconstructed.

## 5.5 Concluding remarks

In this Chapter, a novel methodology has been developed for modelling the existence of a trade relation between any two countries involved in the international

trade of agricultural goods, solely based on the country attributes and link characteristics (such as population, gross domestic product, geographical distance, etc.), without any knowledge about the network structure and topology. More explicitly, applying the methodology to the virtual water trade network, we established that a link is expected to exist depending on the modelled virtual water volume that could be traded along it: a link is considered active when the traded volume is higher than 1000 $m^3$/y, non-active otherwise.
We found that the model is able to correctly predict 84% of the existing links observed in year 2011 (16% of the existing links are "missed"), and 93% of the non-existing ones (7% of the non-existing links are "spurious"). Although the number of missed links might seem large, the associated virtual water flow is negligible (0.1% of the global flow); the links lost by the model, in fact, are mainly those where a minimum volume of virtual water is exchanged.
In order to select the drivers of the link activation, we have performed a stepwise regression; we found that geographical distances between countries, population, and fertilizer use are the fundamental drivers of link activation/deactivation. Binary information (e.g., common languages, common religions) and regional trade agreements are also important, while the other variables, pertaining economical and productive aspects, do not give a significant contribution to the link prediction. These findings extend those of [132] and confirm the importance of geographical distances between countries and population in driving not only the intensity of virtual water flows between pair of countries but also the presence of the link itself.
In the context of analysing the implication of international trade on water resources, the proposed methodology can be applied to predict future evolution of the VWT network taking advantage of the projection already existing for nodes' attributes, such as the gross domestic product [28], the population [115], nitrogen inputs [51] and land-use scenarios [76]. Moreover, the proposed methodology is intended to assess the structure of any weighted network, for which it is possible to get estimates of the link weights only using the node and link characteristics. For instance, the same methodology can be applied to the international trade weighted by different units, such as the food tonnage, the dietary energy equivalent, or the economic value of goods. Although detached from the reality of international trade in a sense (i.e., countries exchange food), all these weights can be exploited to interpret food exchanges. Another possible

application of the link prediction methodology is the backward reconstruction of the bilateral trade matrix before 1986. In fact, the FAO database provides country import and export data since 1961, but bilateral trade flows are not available over the period 1961-1985. The analysis of trade evolution over this period may give important insights given that that these years testify a quick increase of crop yields, through the process known as Green Revolution.

# Chapter 6

# Conclusion

The present thesis contributes to the literature of water footprint assessment, virtual water trade, and water use sustainability in different ways.
The first part of the thesis (Chapter 2) accomplishes the water footprint assessment for different crops under two water use scenarios (i.e., rainfed and irrigated). To this end, a daily soil water balance model has been developed to track the green and blue crop water use at 5'x5' spatial scale. The model makes use of advanced and up-to-date open-access datasets on climatic parameters, soil properties, agricultural practises, and it is applicable to any crop in any year depending on data availability. Moreover, considering multi-cropping patterns enabled improved computations of crop water use by accounting for the different precipitation rates and evapotranspiration demands across the growing seasons. This is particularly important for wheat, maize, and rice that are cultivated twice a year in many regions. The $CWF$ assessment also includes a first-order sensitivity analysis of the estimates to the input parameters, which was missing in the global studies concerning water footprint accounting. The results of the sensitivity analysis show that wheat is mostly sensitive to the length of the growing season, rice to the reference evapotranspiration, maize and soybean to the crop planting date. $CWF$ values are sensitive to input parameters also depending on the location of cultivated areas. The sensitivity analysis performed in this thesis represent a possible way to identify those parameters that farmers and land managers should modify to improve the water use efficiency in order to diminish the pressure exerted on water resources.
The model developed in the first part of the thesis has also been used to validate

the Fast-Track approach recently adopted in the literature of water footprint to deal with the temporal dimension of water footprint (see Chapter 3). The Fast-Track approach assumes the $CWF$ variability as only influenced by the crop yield variations, while it considers the evapotranspiration demand as a constant. In order to test this assumption, the annual $CWF$ estimates obtained with the FT approach have been compared with those obtained through the $CWF$ model (Chapter 2) run on an annual basis from 1961 to 2014. The comparison with coefficient of determination always higher than 0.9 between the two estimates supports and validates the methodology. Also, the uncertainty analysis shows an average error of 10% in the FT approach with respect to the detailed methodology. Providing that this error is three time lower than the error of the model used to estimate the crop water footprint, the FT approach is validated. Once validated, the FT method has been applied to assess the virtual water volumes embedded in the international trade of wheat, rice, maize, and soybean along the period 1986-2011.

The second part of the thesis (Chapter 4) puts the crop water footprint in the context of water use sustainability by comparing it to the local water resources, available from soil moisture, surface water bodies and aquifers. To do so, the water footprint assessment accomplished in Chapter 2 has been extended to other significant crops (i.e., barley, potatoes, sugar cane, sugar beet, and cotton) and the blue water footprint has been computed separated and split into a surface- and ground-water footprint. Such separation is important as surface and ground water resources are differently available across regions and exhibit highly different renewability rates. In order to evaluate the impact of water use on water resources, the green, surface, and ground water footprints have been compared with the renewable water availability, namely water resources annually available from precipitation at the location of crop production. As mentioned earlier in the thesis, literature counts a remarkable variety of contributions that assess agriculture-induced water stress by comparing water consumption to water availability. However, these indicators generally consider all crops together, they do not separate surface water from ground-water, and they often lack a physical interpretation, preventing critical analyses beyond comparisons. The introduction of the Water Debt concept as a new measure of water use sustainability, aims at enriching tools for impact assessment. The Water Debt quantifies the payback time the hydrological cycle takes to replen-

ish the water resources: water debts higher than one year point out threaten regions, where water use critically exceeds water availability. Results show an overall sustainability of green water resources, only some intensively cultivated regions show a water debt of 6-8 months indicating the period of the year during which soil moisture is not available for, e.g., forestry, grazing lands and terrestrial ecosystems. Surface -and ground-water resources appear to be unevenly over-exploited worldwide with water debts ranging from 1 year to over 20 years, underpinning areas of higher impacts, such as over the California Central Valley and the High plain aquifers in the US, the Indo-Gangetic plain between India and Pakistan, in the Amu and Syr Daria basins, and in the Indus river basin.
The last part of this thesis (Chapter 5) focuses on the temporal dynamics of virtual water trade, and in particular on the reconstruction of the network topology. Topology is, in fact, significantly unstable with nearly 30% of links appearing and disappearing from the network year by year. In order to assess the link activation/deactivation, a threshold-based algorithm has been proposed. Differently from the most common link prediction methodologies, the proposed method grounds the link existence on the link weight (i.e., traded volume) estimated as a function of countries properties. Weight estimates have been obtained through a multi-linear regression model that makes use of 20 different regressors. All the links with an estimated weight greater than 1000 $m^3$ are labelled as active links, and non-active otherwise. The algorithm only misses 16% of the existing links and it spuriously adds to the estimated network 7% of non-existing links. Hence, the methodology has proven to be suitable for the objective. Moreover, it has been shown that geographical distances between countries, country population and fertilizers use are the most important drivers of the network topology. Multivariate-linear regressions have also been used to get estimates of the link weights once the topology was known.

***Limitations and future outlook*** The results presented in this thesis are derived on the basis of a number of statistics, input data, and model assumptions. Source of uncertainty are related to the climatic data used to compute the crop water footprint. These data, in fact, are available on 10'x10' grids obtained by *New et al.* [98] through a spatial interpolation of station monthly measures. Given that the model works on a daily scale, a

linear interpolation of the monthly data has been performed. However, such interpolation may hide fluctuations of evapotranspiration demand and extreme rainfall events. Both the spatial and temporal resolution could be improved in future studies by, e.g., integrating global data with local measures and remote sensing information increasingly available, and through the adoption of daily climatic data in the regions where they are available. Another limitation in the water footprint assessment relies in the hypothesis of optimal irrigation timing and quantity every time rainfall is not sufficient to meet the crop water demand (in all the areas equipped for irrigation). Such assumption may lead to overestimate irrigation volumes as blue water may be not available at the site/time of need. Future studies should consider different irrigation scenarios and techniques in order to explore both the sensitivity of crop water footprint to irrigation availability and, particularly, to produce estimates that are as close as possible to the reality.

The water use sustainability assessment provided in the context of this thesis is intended to be a first step in the direction of quantify agricultural impacts on water resources. Specifically, the idea of measuring the sustainability through the payback time is twofold. On the one hand, this measure gives a physical meaning to the more widespread water scarcity indexes, on the other hand it may help connecting and integrating water resources management with other environmental issues, such as the carbon footprint. At the present state, the water debt concept exhibits some limitations that need further developments. Limitations concern, specifically, surface water renewability, which is now considered equal to the locally generated runoff, without considering the upstream generated runoff running in the rivers. Other limitations relate to the uncertainty in the water renewability data derived from WaterGAP. Future researches, while expanding the assessment to other crops and including withdrawals by sectors different from the agricultural one, should focus on the water use sustainability of consumption, considering both the production and the trade of products.

Finally, the present thesis also examines the virtual water trade dynamics considering over 300 products all together over 1986-2011. Given that the link prediction methodology has proven to be useful in predicting the network topology, it can be adopted to reconstruct the past virtual water (bilateral) trade to better understand how it was at the beginning when the US created

an overwhelming preponderance in the world agricultural and food production and trade. Moreover, the methodology can also be used to build possible future scenarios of trade and virtual water trade. It would be very interesting also to disaggregate crops across categories or considering single crop by means (e.g., [135]), e.g., of the tools developed to study multilayer networks [14], which have been recently applied to the ecological networks [106, 105].

***Summary*** The present thesis accomplishes different objectives in the context of the "water-food" nexus and the sustainability of water use. The developed framework is inherently connected and allows to analyse the problem of water resources management under three main aspects. First, the water balance model developed to quantify the crop water footprint at a high spatial resolution provides useful insights and strategies to target improvements of the water use efficiency over production sites. The model allows to study the factors impacting water use efficiency, both the climatic and anthropogenic ones, on a daily basis, and splitting the contribution of green-, surface-, and groundwater sources to the crop water requirement. Moreover, this model is adopted to validate a recently developed Fast-Track methodology to deal with the temporal variability of crop water footprint. Accordingly, we demonstrate how the crop yield is the leading factor of the water footprint temporal evolution. The second important aspect of water management is linked to the global inter-dependence of water resources, which happens through the international trade of agricultural goods. Analysing the implications of trade for water resources is essential to properly manage local water resource and to adopt import strategies when water resources are not sufficient to, e.g., grow food. The novel link prediction algorithm explores the reasons why a link begins or ceases to exist, as a function of, e.g., food demand, land availability, nutrient inputs, gross domestic product. This algorithm allows one to insert the local water management into a more global context. The third aspect focuses on the sustainability of water use, which is a function of both water consumption and local availability. Differently from previous studies, which deal with water use over water availability metrics, with the introduction of the water debt repayment time we try to give a more physical shape to the mismatch between water use and supply. This aspect of the water management is directly linked to both the first and the second aspect of water resource management. On the one hand, it can help prioritizing water efficiency improvement in areas with

scarcer resources; on the other hand, it can be useful to better project crop production intensity and, in some cases, to limit the export of local products to mitigate water stress, or in other cases even to stimulate imports from other regions.

# References


[1] Major river basins of the world, author=Federal Institute of Hydrology , year=2007,.

[2] J. Alcamo, P. Döll, T. Henrichs, F. Kaspar, B. Lehner, T. Rösch, and S. Siebert. Global estimates of water withdrawals and availability under current and future "business-as-usual" conditions. *Hydrological Sciences Journal*, 48(3):339–348, 2003.

[3] J. Alcamo, M. Flörke, and M. Märker. Future long-term changes in global water resources driven by socio-economic and climatic changes. *Hydrological Sciences Journal*, 52(2):247–275, 2007.

[4] J. Alcamo, T. Henrichs, and T. Rosch. World water in 2025. *World Water Series Report*, 2, 2000.

[5] M. Aldaya, A. Chapagain, A. Hoekstra, and M. Mekonnen. *The Water Footprint Assessment Manual: Setting the Global Standard*. Taylor & Francis, 2012.

[6] J. A. Allan. Virtual water-the water, food, and trade nexus. useful concept or misleading metaphor? *Water International*, 28(1):106–113, 2003.

[7] J. A. Allan and T. Allan. *The Middle East water question: Hydropolitics and the global economy*, volume 2. Ib Tauris, 2002.

[8] R. G. Allen, L. Pereira, D. Raes, and M. Smith. Fao irrigation and drainage paper no. 56. *Rome: Food and Agriculture Organization of the United Nations*, pages 26–40, 1998.

[9] J. E. Anderson. A theoretical foundation for the gravity equation. *The American Economic Review*, 69(1):106–116, 1979.

[10] M. Antonelli and M. Sartori. Unfolding the potential of the virtual water concept. what is still under debate? *Environmental Science & Policy*, 2014.

[11] U. Assembly. Transforming our world: the 2030 agenda for sustainable development. resolution adopted by the general assembly on 25 september 2015. new york: United nations; 2015, 2015.

[12] J. H. Bergstrand. The gravity equation in international trade: some microeconomic foundations and empirical evidence. *The review of economics and statistics*, pages 474–481, 1985.

[13] A. Bhaduri, J. Bogardi, A. Siddiqi, H. Voigt, C. Vörösmarty, C. Pahl-Wostl, S. E. Bunn, P. Shrivastava, R. Lawford, S. Foster, et al. Achieving sustainable development goals from a water perspective. *Frontiers in Environmental Science*, 4:64, 2016.

[14] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.

[15] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.

[16] D. Bocchiola, E. Nana, and A. Soncini. Impact of climate change scenarios on crop yield and water footprint of maize in the po valley of italy. *Agricultural Water Management*, 116:50–61, 2013.

[17] A.-M. Boulay, C. Bouchard, C. Bulle, L. Deschênes, and M. Margni. Categorizing water for lca inventory. *The International Journal of Life Cycle Assessment*, 16(7):639–651, 2011.

[18] B. Bouman, E. Humphreys, T. Tuong, and R. Barker. Rice and water. *Advances in Agronomy*, 92:187–237, 2007.

[19] K. A. Brauman, B. D. Richter, S. Postel, M. Malsy, and M. Flörke. Water depletion: An improved metric for incorporating seasonal and dry-year water scarcity into water risk assessments. *Elem Sci Anth*, 4, 2016.

[20] G. H. Brundtland and M. Khalid. Our common future. *New York*, 1987.

[21] J. Butlin. Our common future. by world commission on environment and development.(london, oxford university press, 1987, pp. 383£ 5.95.), 1989.

[22] J. A. Carr, P. D'Odorico, F. Laio, and L. Ridolfi. On the temporal variability of the virtual water network. *Geophysical research letters*, 39(6):L06404, 2012.

[23] J. A. Carr, P. D'Odorico, F. Laio, and L. Ridolfi. Recent history and geography of virtual water trade. *PloS one*, 8(2):e55825, 2013.

[24] I. Cazcarro, R. Duarte, M. Martín-Retortillo, V. Pinilla, and A. Serrano. How sustainable is the increase in the water footprint of the spanish agricultural sector? a provincial analysis between 1955 and 2005–2010. *Sustainability*, 7(5):5094–5119, 2015.

[25] A. Chapagain and A. Hoekstra. The blue, green and grey water footprint of rice from production and consumption perspectives. *Ecological Economics*, 70(4):749–758, 2011.

[26] A. K. Chapagain, A. Hoekstra, and H. Savenije. Water saving through international trade of agricultural products. *Hydrology and Earth System Sciences Discussions*, 10(3):455–468, 2006.

[27] A. K. Chapagain and A. Y. Hoekstra. The global component of freshwater demand and supply: an assessment of virtual water flows between nations as a result of trade in agricultural and industrial products. *Water international*, 33(1):19–32, 2008.

[28] J. C. Cuaresma. Income projections for climate change research: A framework based on human capital dynamics. *Global Environmental Change*, 42:226–236, 2017.

[29] B. C. Curtis, S. Rajaram, H. G. Macpherson, et al. *Bread wheat: improvement and production.* 2002.

[30] C. Dalin and D. Conway. Water resources transfers through southern african food trade: water efficiency and climate signals. *Environmental Research Letters*, 11(1):015005, 2016.

[31] C. Dalin, M. Konar, N. Hanasaki, A. Rinaldo, and I. Rodriguez-Iturbe. Evolution of the global virtual water trade network. *Proceedings of the National Academy of Sciences*, 109(16):5989–5994, 2012.

[32] C. Dalin, S. Suweis, M. Konar, N. Hanasaki, and I. Rodriguez-Iturbe. Modeling past and future structure of the global virtual water trade network. *Geophysical Research Letters*, 39(24):L24402, 2012.

[33] C. Dalin, Y. Wada, T. Kastner, and M. J. Puma. Groundwater depletion embedded in international food trade. *Nature*, 543(7647):700–704, 2017.

[34] P. D'Odorico, J. A. Carr, F. Laio, L. Ridolfi, and S. Vandoni. Feeding humanity through global food trade. *Earth's Future*, 2(9):458–469, 2014.

[35] P. Doell, H. Mueller Schmied, C. Schuh, F. T. Portmann, and A. Eicker. Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and grace satellites. *Water Resources Research*, 50(7):5698–5720, 2014.

[36] P. Döll and K. Fiedler. Global-scale modeling of groundwater recharge. *Hydrology and Earth System Sciences Discussions*, 12(3):863–885, 2008.

[37] P. Döll, H. Hoffmann-Dobrev, F. T. Portmann, S. Siebert, A. Eicker, M. Rodell, G. Strassberg, and B. Scanlon. Impact of water withdrawals from groundwater and surface water on continental water storage variations. *Journal of Geodynamics*, 59:143–156, 2012.

[38] J. Doorenbos, A. Kassam, and C. Bentvelsen. *Yield response to water*. FAO irrigation and drainage paper. Food and Agriculture Organization of the United Nations, 1979.

[39] R. Duarte, V. Pinilla, and A. Serrano. Understanding agricultural virtual water flows in the world from an economic perspective: A long term study. *Ecological Indicators*, 61:980–990, 2016.

[40] D. Dudgeon, A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, M. L. Stiassny, et al. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews*, 81(2):163–182, 2006.

[41] M. Fader, D. Gerten, M. Thammer, J. Heinke, H. Lotze-Campen, W. Lucht, and W. Cramer. Internal and external green-blue agricultural water footprints of nations, and related water and land savings through trade. *Hydrology and Earth System Sciences*, 15(5):1641, 2011.

[42] G. Fagiolo. The international-trade network: gravity equations and topological properties. *Journal of Economic Interaction and Coordination*, 5(1):1–25, 2010.

[43] M. Falkenmark. Meeting water requirements of an expanding world population. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 352(1356):929–936, 1997.

[44] M. Falkenmark. Competing freshwater and ecological services in the river basin perspective: an expanded conceptual framework. *Water international*, 25(2):172–177, 2000.

[45] M. Falkenmark. Growing water scarcity in agriculture: future challenge to global water security. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371(2002):20120410, 2013.

[46] M. Falkenmark, A. Berntell, A. Jägerskog, J. Lundqvist, M. Matz, and H. Tropp. *On the verge of a new water scarcity: a call for good governance and human ingenuity*. Stockholm International Water Institute (SIWI), 2007.

[47] M. Falkenmark and J. Rockström. *Balancing water for humans and nature: the new approach in ecohydrology*. Earthscan, 2004.

[48] M. Falkenmark and J. Rockström. The new blue and green water paradigm: Breaking new ground for water resources planning and management, 2006.

[49] FAO. The state of the world'land and water resources for food and agriculture. *FAO (Food and Agriculture Organziation of the United Nations), Rome*, 2011.

[50] FAO. *Statistical Yearbook 2012: World Food and Agriculture*. 2012.

[51] FAO. Faostat: http://faostat.fao.org, 2014.

[52] FAO. Geonetwork: http://www.fao.org/geonetwork/srv/en/main.home, 2014.

[53] FAO, IIASA, ISRIC, ISSCAS, and JRC. Harmonized world soil database: http://webarchive.iiasa.ac.at/research/luc/external-world-soil-database/html, 2012.

[54] J. Fargione, J. Hill, D. Tilman, S. Polasky, and P. Hawthorne. Land clearing and the biofuel carbon debt. *Science*, 319(5867):1235–1238, 2008.

[55] A. Fracasso. A gravity model of virtual water trade. *Ecological Economics*, 108:215–228, 2014.

[56] M. U. Fratianni. The gravity equation in international trade. 2007.

[57] H. Friedmann. The political economy of food: a global crisis. *New left review*, (197):29, 1993.

[58] L. Getoor and C. P. Diehl. Link mining: a survey. *ACM SIGKDD Explorations Newsletter*, 7(2):3–12, 2005.

[59] T. Gleeson, Y. Wada, M. F. Bierkens, and L. P. van Beek. Water balance of global aquifers revealed by groundwater footprint. *Nature*, 488(7410):197–200, 2012.

[60] P. H. Gleick. A look at twenty-first century water resources development. *Water International*, 25(1):127–138, 2000.

[61] H. C. J. Godfray, J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin. Food security: the challenge of feeding 9 billion people. *science*, 327(5967):812–818, 2010.

[62] N. Hanasaki, T. Inuzuka, S. Kanae, and T. Oki. An estimation of global virtual water flow and sources of water withdrawal for major crops and livestock products using a global hydrological model. *Journal of Hydrology*, 384(3):232–244, 2010.

[63] N. Hanasaki, S. Kanae, T. Oki, K. Masuda, K. Motoya, N. Shirakawa, Y. Shen, and K. Tanaka. An integrated model for the assessment of global water resources–part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences*, 12(4):1007–1025, 2008.

[64] M. A. Hanjra and M. E. Qureshi. Global water crisis and future food security in an era of climate change. *Food Policy*, 35(5):365–377, 2010.

[65] G. L. Hartman, E. D. West, and T. K. Herman. Crops that feed the world 2. soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Security*, 3(1):5–17, 2011.

[66] S. W. Hermanowicz. Sustainability in water resources management: changes in meaning and perception. *Sustainability science*, 3(2):181–188, 2008.

[67] A. Hoekstra and A. Chapagain. *Globalization of Water: Sharing the Planet's Freshwater Resources*. Wiley, 2011.

[68] A. Y. Hoekstra. A critique on the water-scarcity weighted water footprint in lca. *Ecological indicators*, 66:564–573, 2016.

[69] A. Y. Hoekstra. Water footprint assessment: evolvement of a new research field. *Water Resources Management*, pages 1–21, 2017.

[70] A. Y. Hoekstra, A. K. Chapagain, and P. R. van Oel. Advancing water footprint assessment research: Challenges in monitoring progress towards sustainable development goal 6, 2017.

[71] A. Y. Hoekstra and P. Hung. Virtual water trade. *A quantification of virtual water flows between nations in relation to international crop trade. Value of water research report series*, 11:166, 2002.

[72] A. Y. Hoekstra and P. Q. Hung. Globalisation of water resources: international virtual water flows in relation to crop trade. *Global environmental change*, 15(1):45–56, 2005.

[73] A. Y. Hoekstra and M. M. Mekonnen. The water footprint of humanity. *Proceedings of the national academy of sciences*, 109(9):3232–3237, 2012.

[74] A. Y. Hoekstra, M. M. Mekonnen, A. K. Chapagain, R. E. Mathews, and B. D. Richter. Global monthly water scarcity: blue water footprints versus blue water availability. *PLoS One*, 7(2):e32688, 2012.

[75] A. Y. Hoekstra and T. O. Wiedmann. Humanity's unsustainable environmental footprint. *Science*, 344(6188):1114–1117, 2014.

[76] G. C. Hurtt, L. P. Chini, S. Frolking, R. Betts, J. Feddema, G. Fischer, J. Fisk, K. Hibbard, R. Houghton, A. Janetos, et al. Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic change*, 109(1-2):117, 2011.

[77] Å. Jansson, C. Folke, J. Rockström, L. Gordon, and M. Falkenmark. Linking freshwater flows and ecosystem services appropriated by people: the case of the baltic sea drainage basin. *Ecosystems*, 2(4):351–366, 1999.

[78] W. Jiang and R. Marggraf. Bilateral virtual water trade in agricultural products: a case study of germany and china. *Water International*, 40(3):483–498, 2015.

[79] M. Konar and K. Caylor. Virtual water trade and development in africa. *Hydrology and Earth System Sciences*, 17(10):3969–3982, 2013.

[80] M. Konar, C. Dalin, N. Hanasaki, A. Rinaldo, and I. Rodriguez-Iturbe. Temporal dynamics of blue and green virtual water trade networks. *Water Resources Research*, 48(7):W07509, 2012.

[81] M. Konar, C. Dalin, S. Suweis, N. Hanasaki, A. Rinaldo, and I. Rodriguez-Iturbe. Water for food: The global virtual water trade network. *Water Resources Research*, 47(5):W05520, 2011.

[82] A. Kounina, M. Margni, J.-B. Bayart, A.-M. Boulay, M. Berger, C. Bulle, R. Frischknecht, A. Koehler, L. M. i Canals, M. Motoshita, et al. Review of methods addressing freshwater use in life cycle inventory and impact assessment. *The International Journal of Life Cycle Assessment*, 18(3):707–721, 2013.

[83] M. Kummu, J. Guillaume, H. de Moel, S. Eisner, M. Flörke, M. Porkka, S. Siebert, T. Veldkamp, and P. Ward. The world's road to water scarcity: Shortage and stress in the 20th century and pathways towards sustainability. *Scientific Reports*, 6, 2016.

[84] R. Labrada. The need for improved weed management in rice. fao: Proceedings of the 20th session of the international rice commission (bangkok, thailand, 23–26 july 2002), 2012.

[85] J. Liu and H. Yang. Spatially explicit assessment of global consumptive water uses in cropland: Green and blue water. *Journal of Hydrology*, 384(3):187–197, 2010.

[86] D. B. Lobell, K. G. Cassman, and C. B. Field. Crop yield gaps: their importance, magnitudes, and causes. *Annual Review of Environment and Resources*, 34(1):179, 2009.

[87] L. Lü and T. Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150–1170, 2011.

[88] L. Marston, M. Konar, X. Cai, and T. J. Troy. Virtual groundwater transfers from overexploited aquifers in the united states. *Proceedings of the National Academy of Sciences*, 112(28):8561–8566, 2015.

[89] M. Mekonnen and A. Hoekstra. The green, blue and grey water footprint of crops and derived crop products. *Hydrology & Earth System Sciences Discussions*, 8(1), 2011.

[90] M. Mekonnen and A. Hoekstra. Water footprint benchmark for crop production. 2013.

[91] M. M. Mekonnen and A. Y. Hoekstra. The green, blue and grey water footprint of crops and derived crop products. 2010.

[92] M. M. Mekonnen and A. Y. Hoekstra. National water footprint accounts: the green, blue and grey water footprint of production and consumption. 2011.

[93] M. M. Mekonnen and A. Y. Hoekstra. Four billion people facing severe water scarcity. *Science advances*, 2(2):e1500323, 2016.

[94] C. Monfreda, N. Ramankutty, and J. A. Foley. Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global biogeochemical cycles*, 22(1):GB1022, 2008.

[95] N. D. Mueller, J. S. Gerber, M. Johnston, D. K. Ray, N. Ramankutty, and J. A. Foley. Closing yield gaps through nutrient and water management. *Nature*, 490(7419):254–257, 2012.

[96] H. Müller Schmied, S. Eisner, D. Franz, M. Wattenbach, F. T. Portmann, M. Flörke, and P. Döll. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, 18(9):3511–3538, 2014.

[97] H. A. Munia, J. Guillaume, N. Mirumachi, Y. Wada, and K. Matti. How downstream sub-basins depend on upstream inflows to avoid scarcity: typology and global analysis of transboundary rivers. *Hydrology and Earth System Sciences Discussions*, 2017:1–26, 2017.

[98] M. New, D. Lister, M. Hulme, and I. Makin. A high-resolution data set of surface climate over global land areas. *Climate research*, 21(1):1–25, 2002.

[99] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[100] T. Oki and S. Kanae. Virtual water trade and world water resources. *Water Science and Technology*, 49(7):203–209, 2004.

[101] T. Oki and S. Kanae. Global hydrological cycles and world water resources. *science*, 313(5790):1068–1072, 2006.

[102] H. G. Overman, S. J. Redding, and A. J. Venables. The economic geography of trade production and income: a survey of empirics. 2001.

[103] J.-F. Pekel, A. Cottam, N. Gorelick, A. S. Belward, et al. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633):418–422, 2016.

[104] S. Pfister, P. Bayer, A. Koehler, and S. Hellweg. Environmental impacts of water use in global crop production: hotspots and trade-offs with land use. *Environmental science & technology*, 45(13):5761–5768, 2011.

[105] S. Pilosof, M. A. Porter, M. Pascual, and S. Kéfi. The multilayer nature of ecological networks. *Nature ecology & evolution*, 1:0101, 2017.

[106] M. J. Pocock, D. M. Evans, and J. Memmott. The robustness and restoration of a network of ecological networks. *Science*, 335(6071):973–977, 2012.

[107] M. Porkka, J. H. Guillaume, S. Siebert, S. Schaphoff, and M. Kummu. The use of food imports to overcome local limits to growth. *Earth's Future*, 5(4):393–407, 2017.

[108] F. T. Portmann, S. Siebert, and P. Döll. Mirca2000—global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global biogeochemical cycles*, 24(1):GB1011, 2010.

[109] W. Pute, W. Yubao, Z. Xining, S. Shikun, and J. Jiming. Spatiotemporal variation in water footprint of grain production in china. *Frontiers of Agricultural Science and Engineering*, 2(2):186–193, 2015.

[110] D. K. Ray, J. S. Gerber, G. K. MacDonald, and P. C. West. Climate variation explains a third of global crop yield variability. *Nature communications*, 6, 2015.

[111] D. K. Ray, N. Ramankutty, N. D. Mueller, P. C. West, and J. A. Foley. Recent patterns of crop yield growth and stagnation. *Nature communications*, 3:1293, 2012.

[112] J. Rockström. Water for food and nature in drought–prone tropics: vapour shift in rain–fed agriculture. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 358(1440):1997–2009, 2003.

[113] J. Rockström and J. Barron. Water productivity in rainfed systems: overview of challenges and analysis of opportunities in water scarcity prone savannahs. *Irrigation Science*, 25(3):299–311, 2007.

[114] S. Rost, D. Gerten, A. Bondeau, W. Lucht, J. Rohwer, and S. Schaphoff. Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research*, 44(9), 2008.

[115] K. Samir and W. Lutz. The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100. *Global Environmental Change*, 42:181–192, 2017.

[116] B. R. Scanlon, C. C. Faunt, L. Longuevergne, R. C. Reedy, W. M. Alley, V. L. McGuire, and P. B. McMahon. Groundwater depletion and sustainability of irrigation in the us high plains and central valley. *Proceedings of the national academy of sciences*, 109(24):9320–9325, 2012.

[117] L. Scherer and S. Pfister. Dealing with uncertainty in water scarcity footprints. *Environmental Research Letters*, 11(5):054008, 2016.

[118] J. Schwarz, E. Mathijs, and M. Maertens. Changing patterns of global agri-food trade and the economic efficiency of virtual water flows. *Sustainability*, 7(5):5542–5563, 2015.

[119] P. A. Seck, A. Diagne, S. Mohanty, and M. C. Wopereis. Crops that feed the world 7: Rice. *Food Security*, 4(1):7–24, 2012.

[120] D. Seckler, R. Barker, and U. Amarasinghe. Water scarcity in the twenty-first century. *International Journal of Water Resources Development*, 15(1-2):29–42, 1999.

[121] J. Shi, J. Liu, and L. Pinter. Recent evolution of china's virtual water trade: analysis of selected crops and considerations for policy. *Hydrology and Earth System Sciences*, 18(4):1349–1357, 2014.

[122] J. Shi, Z. Wang, Z. Zhang, Y. Fei, Y. Li, F. Zhang, J. Chen, and Y. Qian. Assessment of deep groundwater over-exploitation in the north china plain. *Geoscience Frontiers*, 2(4):593–598, 2011.

[123] B. Shiferaw, B. M. Prasanna, J. Hellin, and M. Bänziger. Crops that feed the world 6. past successes and future challenges to the role played by maize in global food security. *Food Security*, 3(3):307–327, 2011.

[124] B. Shiferaw, M. Smale, H.-J. Braun, E. Duveiller, M. Reynolds, and G. Muricho. Crops that feed the world 10. past successes and future challenges to the role played by wheat in global food security. *Food Security*, 5(3):291–317, 2013.

[125] S. Siebert, J. Burke, J.-M. Faures, K. Frenken, J. Hoogeveen, P. Döll, and F. T. Portmann. Groundwater use for irrigation–a global inventory. *Hydrology and Earth System Sciences*, 14(10):1863–1880, 2010.

[126] S. Siebert and P. Döll. Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. *Journal of Hydrology*, 384(3):198–217, 2010.

[127] I. Soligno, L. Ridolfi, and F. Laio. The environmental cost of a reference withdrawal from surface waters: Definition and geography. *Advances in Water Resources*, 110:228–237, 2017.

[128] S. Sun, P. Wu, Y. Wang, and X. Zhao. Temporal variability of water footprint for maize production: The case of beijing from 1978 to 2008. *Water resources management*, 27(7):2447–2463, 2013.

[129] S. Sun, P. Wu, Y. Wang, X. Zhao, J. Liu, and X. Zhang. The impacts of interannual climate variability and agricultural inputs on water footprint of crop production in an irrigation district of china. *Science of the Total Environment*, 444:498–507, 2013.

[130] S. Suweis, M. Konar, C. Dalin, N. Hanasaki, A. Rinaldo, and I. Rodriguez-Iturbe. Structure and controls of the global virtual water trade network. *Geophysical Research Letters*, 38(10), 2011.

[131] S. Tamea, P. Allamano, J. Carr, P. Claps, F. Laio, and L. Ridolfi. Local and global perspectives on the virtual water trade. *Hydrology and Earth System Sciences*, 17(3):1205–1215, 2013.

[132] S. Tamea, J. Carr, F. Laio, and L. Ridolfi. Drivers of the virtual water trade. *Water Resources Research*, 50(1):17–28, 2014.

[133] D. Tilman, C. Balzer, J. Hill, and B. L. Befort. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, 108(50):20260–20264, 2011.

[134] D. Tilman, K. G. Cassman, P. A. Matson, R. Naylor, and S. Polasky. Agricultural sustainability and intensive production practices. *Nature*, 418(6898):671–677, 2002.

[135] S. Torreggiani, G. Mangioni, M. J. Puma, and G. Fagiolo. Identifying the community structure of the international food-trade multi network. *arXiv preprint arXiv:1711.05784*, 2017.

[136] M. Tuninetti, S. Tamea, P. D'Odorico, F. Laio, and L. Ridolfi. Global sensitivity of high-resolution estimates of crop water footprint. *Water Resources Research*, 51(10):8257–8272, 2015.

[137] M. Tuninetti, S. Tamea, F. Laio, and L. Ridolfi. To trade or not to trade: Link prediction in the virtual water network. *Advances in Water Resources*, pages 1–10, 2016.

[138] M. Tuninetti, S. Tamea, F. Laio, and L. Ridolfi. A fast track approach to deal with the temporal dimension of crop water footprint. *Environmental Research Letters*, 12(7):074010, 2017.

[139] p. d. UN. World population prospects the 2012 revision. 2013.

[140] UniversityofEastAnglia. Cru: http://www.cru.uea.ac.uk/data, 2014.

[141] D. Vanham, A. Y. Hoekstra, Y. Wada, F. Bouraoui, A. de Roo, M. Mekonnen, W. van de Bund, O. Batelaan, P. Pavelic, W. Bastiaanssen, et al. Physical water scarcity metrics for monitoring progress towards sdg target 6.4: An evaluation of indicator 6.4. 2 "level of water stress". *Science of the Total Environment*, 613:218–232, 2018.

[142] C. J. Vörösmarty, P. Green, J. Salisbury, and R. B. Lammers. Global water resources: vulnerability from climate change and population growth. *science*, 289(5477):284–288, 2000.

[143] C. J. Vörösmarty, P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. Reidy Liermann, et al. Global threats to human water security and river biodiversity. *nature*, 467(7315):555, 2010.

[144] Y. Wada and M. F. Bierkens. Sustainability of global water use: past reconstruction and future projections. *Environmental Research Letters*, 9(10):104003, 2014.

[145] Y. Wada, I. E. de Graaf, and L. P. van Beek. High-resolution modeling of human and climate impacts on global water resources. *Journal of Advances in Modeling Earth Systems*, 8(2):735–763, 2016.

[146] Y. Wada, M. Flörke, N. Hanasaki, S. Eisner, G. Fischer, S. Tramberend, Y. Satoh, M. Van Vliet, P. Yillia, C. Ringler, et al. Modeling global water use for the 21st century: Water futures and solutions (wfas) initiative and its approaches. *Geoscientific Model Development*, 9:175–222, 2016.

[147] Y. Wada, L. P. van Beek, C. M. van Kempen, J. W. Reckman, S. Vasak, and M. F. Bierkens. Global depletion of groundwater resources. *Geophysical Research Letters*, 37(20), 2010.

[148] T. Wickham and V. Singh. Water movement through wet soils. *Soils and rice*, pages 337–358, 1978.

[149] Y. Xu, K. Huang, Y. Yu, and X. Wang. Changes in water footprint of crop production in beijing from 1978 to 2012: a logarithmic mean divisia index decomposition analysis. *Journal of Cleaner Production*, 87:180–187, 2015.

[150] H. Yang, L. Wang, K. C. Abbaspour, and A. J. Zehnder. Virtual water trade: an assessment of water use efficiency in the international food trade. *Hydrology and Earth System Sciences Discussions*, 10(3):443–454, 2006.

[151] S. Yano, N. Hanasaki, N. Itsubo, and T. Oki. Water scarcity footprints by considering the differences in water sources. *Sustainability*, 7(8):9753–9772, 2015.

[152] S. Yano, N. Hanasaki, N. Itsubo, and T. Oki. Potential impacts of food production on freshwater availability considering water sources. *Water*, 8(4):163, 2016.

[153] L. Zhuo, M. Mekonnen, and A. Hoekstra. Sensitivity and uncertainty in crop water footprint accounting: a case study for the yellow river basin. *Hydrology and Earth System Sciences Discussions*, 11(1):135–167, 2014.

[154] L. Zhuo, M. M. Mekonnen, and A. Y. Hoekstra. The effect of inter-annual variability of consumption, production, trade and climate on crop-related green and blue water footprints and inter-regional virtual water trade: A study for china (1978–2008). *Water research*, 94:73–85, 2016.

[155] L. Zhuo, M. M. Mekonnen, A. Y. Hoekstra, and Y. Wada. Inter-and intra-annual variation of water footprint of crops and blue water scarcity in the yellow river basin (1961–2009). *Advances in Water Resources*, 87:29–41, 2016.